1-1-2014

Melody as Prosody: Toward a Usage-Based Theory of Music

Thomas Mathew Pooley
University of Pennsylvania, tompooly@gmail.com

Melody as Prosody: Toward a Usage-Based Theory of Music

Abstract
MELODY AS PROSODY: TOWARD A USAGE-BASED THEORY OF MUSIC

Thomas M. Pooley

Gary A. Tomlinson

Rationalist modes of inquiry have dominated the cognitive science of music over the past several decades. This dissertation contests many rationalist assumptions, including its core tenets of nativism, modularity, and computationism, by drawing on a wide range of evidence from psychology, neuroscience, linguistics, and cognitive music theory, as well as original data from a case study of Zulu song prosody. An alternative biocultural approach to the study of music and mind is outlined that takes account of musical diversity by attending to shared cognitive mechanisms. Grammar emerges through use, and cognitive categories are learned and constructed in particular social contexts. This usage-based theory of music shows how domain-general cognitive mechanisms for patterning-finding and intention-reading are crucial to acquisition, and how Gestalt principles are invoked in perception. Unlike generative and other rationalist approaches that focus on a series of idealizations, and the cognitive `competences' codified in texts and musical scores, the usage-based approach investigates actual performances in everyday contexts by using instrumental measures of process.

The study focuses on song melody because it is a property of all known musics. Melody is used for communicative purposes in both song and speech. Vocalized pitch patterning conveys a wide range of affective, propositional, and syntactic information through prosodic features that are shared by the two domains. The study of melody as prosody shows how gradient pitch features are crucial to the design and communicative functions of song melodies. The prosodic features shared by song and speech include: speech tone, intonation, and pitch-accent. A case study of ten Zulu memulo songs shows that pitch is not used in the discrete or contrastive fashion proposed by many cognitive music theorists and most (generative) phonologists. Instead there are a range of pitch categories that include pitch targets, glides, and contours. These analyses also show that song melody has a multi-dimensional pitch structure, and that it is a dynamic adaptive system that is irreducible in its complexity.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Music

First Advisor
Gary A. Tomlinson

Keywords
melody, music cognition, prosody, song, speech, usage-based theory

Subject Categories
Linguistics | Music | Psychology

MELODY AS PROSODY:
TOWARD A USAGE-BASED THEORY OF MUSIC

Thomas Mathew Pooley

A DISSERTATION

in

Music

Presented to the Faculties of the University of Pennsylvania

In Partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

2014

Supervisor of Dissertation

Gary A. Tomlinson
John Hay Whitney Professor of Music and Humanities, Yale University

Graduate Group Chairperson

Jairo Moreno
Associate Professor of Music

Dissertation Committee:

Carol A. Muller
Professor of Music

Eugene Narmour
Emeritus Edmund J. Kahn Distinguished Professor of Music

Dedication

To My Mother, Elsa Susanna Pooley

Acknowledgment

My very first seminar at Penn was 'biocultural musicology' with Gary Tomlinson. This proved to be a major turning point in my academic career because it introduced me to an entirely new world of scholarship in the cognitive sciences. I soon set out on a course of study that included seminars in the departments of anthropology, biology, linguistics, philosophy, psychology, and statistics, in addition to the core offerings in music. Gary supported all this and he became my mentor and the supervisor of this dissertation. His formidable command of literatures in many interrelated disciplines, and his sheer enthusiasm for scientific pursuits in the humanities, emboldened me to plot a wide interdisciplinary course. I am most grateful for his meticulous reading of this manuscript and the many improvements to the prose and argument that resulted.

A second turning point was an unexpected graduate seminar with Eugene Narmour. He also guided me in the study of cognitive music theory, and he introduced me to the work of the Penn school in particular. Our many discussions out at St Joseph's Way shaped my thinking on diverse topics. I am sincerely grateful for his mentorship and support, his eagle eye for detail, and his careful scrutiny of this manuscript. Kathy and Gene were always generous and gracious hosts during my occasional visits to Philadelphia, and I have thoroughly enjoyed their company and friendship.

I was privileged to study with several other influential teachers in the Penn music department. Carol Muller introduced me to Philadelphia and the United States as a fellow South African, and to the exciting new realm of online learning. Timothy Rommen was an important sounding board as I began to explore literatures in the history of ethnomusicology. His seminar on fieldwork was vital preparation for the case study described here. Thank you also to the other extraordinary faculty members with whom I studied, including: Jeff Kallberg, Jairo Moreno, Carolyn Abbate, Emma Dillon, Emily Dolan, Jim Primosch, Guy Ramsey, and Anna Weesner.

Several other faculty at Penn provided important intellectual support and guidance. Theodore Schurr in the anthropology department taught a comprehensive seminar in physical anthropology, and advised me on further studies in evolutionary anthropology. Robert Seyfarth opened doors in the psychology department so that I could study cognitive neuroscience and the biological basis of behavior. Lori Flanagan-Cato introduced me to neuroscience with a baptism by fire (and it worked!). In linguistics, Jiahong Yuan introduced me to phonetics and speech prosody with patience and encouragement, and Rolf Noyer kindly allowed me to audit his graduate seminar in generative phonology. Michael Weisberg convened a brilliant seminar on the philosophy of biology, and challenged me to refocus my thinking on the evolution of music and language. All of these diverse perspectives prepared the ground for the arguments set out in this dissertation.

My doctoral studies were generously funded by a Benjamin Franklin Fellowship from the University of Pennsylvania, a fellowship from the Booth Ferris Foundation, a Cross Travel Grant, a Penfield Research Award, and study assistance from the University of South Africa. Without the financial support of these institutions and organizations I would not have been able to study at Penn or to conduct any of the research upon which this dissertation is based. My thanks to each of them.

Fieldwork is the art of collaboration and I was fortunate to have many collaborators. I learned an enormous amount through conversations with my *umfowethu* Thulani 'Roger' Gumede. Thank you also to my guides, translators, and drivers: Ndididi Dladla, Elliot Banda, Caiphus Mthabela, Zamani Madonsela, and Zama Nzuza. I would like to thank the following individuals, families, groups, and institutions for their diverse contributions to the 'Zulu Song Project': GG Alcock, Bergville Blue Roses, Eric Biyela, Cavern Choir, Entokozweni Primary School, Eshowe High School Choir, Busisiwe Gumede, Phumzile Gumede, Rose Hlophe, Himeville Anglican Church, Impolimpoli Primary School, Mvezi Khumalo, Mafutha Enkukhu, Izainah Masoka, Mbhele Family, Inkosi Miya, Vincent Mbatha, Wiseman Mbatha, Caiphus Mthabela, Jabulani Mthembu, JJ Mthethwa, Muthi Futhi, Menzi Ngubane, Sula Ngwenya, Nokulunga, Nomusa Nhlapho, Gana Nyawose, Brother Clement Sithole, St Philips Primary School, Udumo Gospel Choir, Pessa Weinberg, Gill Whittington Banda, and Sithembile Xulu.

The Cavern Drakensberg Resort and Spa deserves as a special word of thanks. To Megan Carte Bedingham, Hilton Bedingham, Leslie Carte, Murray Carte, Peter Carte, and Rhona Carte: thank you for your friendship, hospitality, expert advice, and homely support—all in a spectacular setting in the Drakensberg mountains. Thank you also to Ezemvelo KwaZulu-Natal Wildlife, Ndumo Game Reserve, for providing me with research accommodation in April 2012. To Anne Rennie, with whom I enjoyed watching many a famous tennis match, thank you for hosting me at Sunset. Rauri Alcock shared valuable audio-visual recordings with me, and went to extraordinary efforts to introduce me to communities in the Msinga and Weenen districts. His advice was critical to my negotiating the fieldwork process, and to drawing up the informed consent documents. I have many happy memories of Mdukatshani thanks to the hospitality and friendship of Creina Alcock. Her insight into rural life, and her knowledge of archival and oral histories are unrivalled in scope and detail. Our conversations will not come to an end.

My family in Georgetown, including James, Christine, William, Dorothee, and Victoria, proved to be close indeed. William was an excellent tour guide during my first summer in DC, and he continued to ask probing questions about my work (even when I was trying to take a holiday from it!). James and Christine hosted me on innumerable occasions and provided the home away from home that I needed. In Cape Town, William and Winifred Bond were kind hosts and inquiring interlocutors, and I enjoyed the cross-disciplinary pollinations that resulted.

Thank you to my dear friends Peter and Dayna Mondelli, and Myron Gray, for many good times in Philadelphia. Karin Meyer was my companion through much of this journey, and her friendship and support were very special to me. My brothers Simon and Justin encouraged me from the first. Their friendship and counsel offered sure solace in periods of uncertainty, and shared joy in times of excitement and triumph. Finally, during the many years away, and the months and years back home, it was my mother who encouraged, supported, and reminded me, more than anyone else, of the convictions that guided me in this project and which brought it to a conclusion: I dedicate this dissertation to her, with love and gratitude.

ABSTRACT

MELODY AS PROSODY:

TOWARD A USAGE-BASED THEORY OF MUSIC

Thomas M. Pooley

Gary A. Tomlinson

Rationalist modes of inquiry have dominated the cognitive science of music over the past several decades. This dissertation contests many rationalist assumptions, including its core tenets of nativism, modularity, and computationism, by drawing on a wide range of evidence from psychology, neuroscience, linguistics, and cognitive music theory, as well as original data from a case study of Zulu song prosody. An alternative biocultural approach to the study of music and mind is outlined that takes account of musical diversity by attending to shared cognitive mechanisms. Grammar emerges through use, and cognitive categories are learned and constructed in particular social contexts. This usage-based theory of music shows how domain-general cognitive mechanisms for patterning-finding and intention-reading are crucial to acquisition, and how Gestalt principles are invoked in perception. Unlike generative and other rationalist approaches that focus on a series of idealizations, and the cognitive `competences' codified in texts and musical scores, the usage-based approach investigates actual performances in everyday contexts by using instrumental measures of process.

The study focuses on song melody because it is a property of all known musics. Melody is used for communicative purposes in both song and speech. Vocalized pitch patterning conveys a wide range of affective, propositional, and syntactic information through prosodic features that are shared by the two domains. The study of melody as prosody shows how gradient pitch features are crucial to the design and communicative functions of song melodies. The prosodic features shared by song and speech include: speech tone, intonation, and pitch-accent. A case study of ten Zulu memulo songs shows that pitch is not used in the discrete or contrastive fashion proposed by many cognitive music theorists and most (generative) phonologists. Instead there are a range of pitch categories that include pitch targets, glides, and contours. These analyses also show that song melody has a multi-dimensional pitch structure, and that it is a dynamic adaptive system that is irreducible in its complexity.

Table of Contents

List of illustrations:

List of musical examples:

CHAPTER 1

Biocultural foundations of music and mind

The capacity for song is a species-specific trait that distinguishes us from all other primates. It manifests itself in early childhood and its forms and functions multiply thereafter, assuming a bewildering array of cultural instantiations. Yet despite this diversity there is one distinctive feature that is common across all cultures: melody. All songs possess a distinctive pitch pattern or melody. This is an undisputed and universal feature of song and, to some degree, of linguistic utterance too. Melody is an intrinsic property of vocalization, since air channeled through the vocal tract causes periodic vibrations of the glottis that are transmitted as information-rich sound waves to the speaker and listener. The frequencies of these sound waves are filtered through the auditory system, encoded, and interpreted as salient social and environmental information. Humans have adapted this melodic capacity for many basic communicative functions. For instance, all speakers rely on intonation and pitch-accent to deliver propositional and affective information that directs the attention of listeners. In addition, most of the world's languages use semantic tone to disambiguate words. Mothers use a form of heightened intonation or infant-directed speech to communicate with their children. And song itself is an integral part of many cultural practices from children's rounds to religious rites. These are all forms of melody that give meaning to our shared sense of humanity and that allow us to communicate with precision and purpose in daily life. There is even evidence that this capacity has a genetic basis, a perspective that reinforces the sense in which it is basic to all humans and one that suggests a deep history.[1]

The study of song melody is complex because it is so multi-dimensional. Pitch patterning in song serves several musical and linguistic functions simultaneously. One important goal of this dissertation is to draw attention to these diverse functions and to develop a framework for establishing their interrelationships. Prior studies have tended to underestimate this complexity by focusing on the parameters prescribed by Western staff notation. This ethnocentrism has meant that the extraordinary diversity of pitch patterning in song has been dramatically underestimated. One fundamental oversight has been to ignore suprasegmental prosodic features and elements of gradience that characterize all vocalization. The case study of Zulu song presented in Chapter 4 is a first step in addressing this gap in knowledge.

To accomplish these goals this study develops a new *usage-based* framework for studying melody as a *complex adaptive system*. The first three chapters of the dissertation outline the theoretical basis for this approach and offer a critique of the current state of research. Since the 1980s, cognitive music theory has been dominated by generative and other formalist modes of inquiry that assume a rationalist, nativist account of human musicking. In Chapters 1 and 2 I show how much recent evidence from across the cognitive sciences does not support the basic tenets of this view. In contrast, the usage-

---

[1] Tone-deaf individuals constitute approximately two to three percent of the population, suggesting that

based framework focuses on *production* rather than on listener competence and draws evidence from actual performances rather than musical scores and notated transcriptions. It is an approach that builds on recent work in linguistics and cognitive science that takes higher cognitive functions to be the product of a set of learning mechanisms that are domain-general.[2] To resolve problems in the study of music mind we need a bottom-up, data-driven approach grounded in empirical investigations of music in world cultures.

There are several advantages to usage-based accounts. One is the emphasis on the *constructive* role of learning. Unlike nativism, in which the focus is on the competence of 'experienced listeners' of musical works, this approach is based on studies of actual performances. The premise is that any theory of song must account for elements of *production* rather than only perception since the variation and gradience introduced by usage results in new modes of communication; indeed, it is through the complexity of social interactions that new cultural conventions emerge and that grammaticalization takes place.[3] Nativist theories are simply too rigid to account for the dynamics of this system. Rather than assume song to be determined by a set of pre-wired competencies, I

[2] The usage-based approach is outlined clearly and succinctly in Michael Tomasello's book, *Constructing a Language: A Usage-Based Theory of Language Acquisition* (Cambridge, MA: Harvard University Press, 2003). This text will be discussed in more detail in Part 3 of Chapter 2.

[3] 'Grammaticalization' refers to the historical process by which humans have developed the ability to create grammars, both musical and linguistic. "Human beings use their linguistic symbols together in patterned ways, and these patterns, known as linguistic constructions, take on meanings of their own—deriving partly from the meanings of the individual symbols but, over time, at least partly from the pattern itself" (Tomasello, *Constructing a Language*, 8). Given that this process of grammaticalization relies on "domain-general cognitive and social-cognitive processes that operate as people communicate with and learn from one another" (Ibid.), we may hypothesize that music relies on the same, or similar, mechanisms.

approach song as an *open system* in constant flux. There are several variables that interact and introduce feed-back and feed-forward loops in a state of dynamic equilibrium.

The relationship between song and speech, music and language is complex; songs invariably have words and singing employs the same vocal apparatus as does speech. For these reasons I treat song and speech as integrated systems and draw on linguistic and music theory in equal measure in my analyses. This chapter provides an introduction to many of the issues and ideas that will be raised later on in the dissertation, and it focuses on important questions of specificity. It is divided into two main parts: the first surveys research on the underlying mechanisms for the acquisition and learning of song, and the second investigates the nature of musical diversity and universals.

PART 1: MECHANISMS

1.1 Innateness and acquisition

It is important to understand not only how, when, and in what ways infants acquire the ability to sing, but also the extent to which this ability is biologically encoded. How does this compare with language? The acquisition of these behaviors through childhood reflects not only the distinctiveness of the two domains of music and language, and what is special to each, but also the extent to which the elements of melody are *learned*. This issue is however a contested one in cognitive science. Nativists and non-nativists disagree over the relative importance of the human cognitive endowment in accounting for these abilities. Nativists assert that there are domain-specific mechanisms encoded in the genome that determine music and linguistic behaviors, and these cannot be learned through ontogeny. Chomsky's well-known argument for the 'poverty of the stimulus,' for instance, claims that 'knowledge of language' is too rich to be acquired through experience and must therefore be innately acquired (see Chapter 2 for a more detailed discussion of his position). Theories of nativism play a central theoretical role in studies of generative grammar and language in general, as Ray Jackendoff points out:

> A major dispute in the theory of language, of course, is how much of the language acquisition capacity is special-purpose. Many people (e.g., Christiansen & Chater, 2008; Tomasello, 2003) think that language is acquired through general-purpose learning plus abilities for social interaction. This view is explicitly in opposition to the claims of generative grammar up to the early 1990s to the effect that there must be a rich innate language-specific Universal Grammar (Chomsky, 1965, 1981). In between these two extremes are all manner of intermediate views (Hauser, Chomsky, & Fitch, 2002; Jackendoff, 2002; Pinker & Jackendoff, 2005). […] There seems to be less vehement dispute about the parallel issue in music cognition and acquisition, perhaps only because claims for an innate music capacity have been less highly politicized—but perhaps also because claims that music is an adaptation favored by natural selection are considerably weaker than those for language.[4]

The disputes that have arisen in linguistics as a result of these contrasting positions have seen no parallel in music studies. Most music cognitivists today assume a nativist position despite evidence to the contrary. The purpose of this chapter and the next is to

[4] Ray Jackendoff, "Parallels and Nonparallels between Language and Music," *Music Perception* 26, No. 3 (February 2009): 196.

draw attention to this evidence, to contest many of the theoretical assumptions of the nativist position, and then to offer an alternative.

Nativists and non-nativists disagree over whether or not innate mechanisms account for the acquisition of music and language as specialized behaviors. Non-nativists—who, for convenience sake, I will describe as *empiricists* in this dissertation—claim that a small number of domain-general learning mechanisms and a rich social-cognitive environment explain acquisition. By contrast, nativists propose an 'innate' 'language acquisition device' that is encoded in the genome.

> The now well-known idea put forth by Chomsky (1965) was that humans enter the world equipped with a "language acquisition device" (LAD). That is, given proper input, we have the cognitive equipment to automatically acquire language. Because of the presence of precious perceptual abilities of infants, a nativist position is held by the majority of researchers in this field. Our intrinsic capacity for music is seen as leaning to nature, while the learning of specific musical forms relies on nurture—specifically on exposure during a sensitive or critical period of development (Trainor 2005). This strong nativist position in the literature on early musical capacities suggests that a counterpart in music to the LAD indeed exists, which we will call the "music acquisition device."[5]

Is it not possible that the LAD and MAD (music acquisition device) are one and the same thing, or that there are in fact more general cognitive mechanisms underlying complex

[5] Daniel Levitin and Anna Tirovolas, "Current advances in the cognitive neuroscience of music," *Annals of the New York Academy of Sciences* 1156, Iss. 1 (March 2009): 216. See the discussion of nativism in Chapter 2 for more details on Chomsky's work.

communicative behaviors? A usage-based account points to such domain-general mechanisms to show that it is not only the "specific musical" forms that are learned, but also the (grammatical) principles by which they are constructed. Given the plasticity of the human infant mind/brain we need to be cautious in adopting too strong a nativist hypothesis. Evidence from studies of language acquisition, infant-directed speech, and music in early childhood development show that music and language share important prosodic features from the first, and that these are only later recruited for more specialized communicative ends. Section 1.2 surveys this important literature.

1.2 Infant-Directed Speech

The roots of *prosodic* structures are evident early in childhood and play an important role in the acquisition of language and music, too. Within the first six months to a year of infancy mothers communicate with their infants using a form of 'infant-directed speech' (IDS) and many features of IDS are important to the development of both adult song and speech. Sandra Trehub describes the musical characteristics of IDS, in relation to adult speech, as follows: IDS has higher pitch height, slower tempo, highly regular timing, and exaggerated expressive variations in dynamics.[6] Infants respond better to IDS than to adult speech and it has been observed across many cultures as the primary mode of interaction between caregivers and infants. Studies show that infants "are predisposed to selectively attend to the distinctive pitch contours of infant-directed speech, whose

[6] Sandra Trehub, "Musical predispositions in infancy," *Annals of the New York Academy of Sciences*, 930, Iss. 1 (2001): 1-16.

primitive emotional meanings can be decoded in the absence of language."[7] These arousal and affective properties derive from prosodic features of IDS, some of which may be carried through into adulthood. IDS may also aid the "perceptual learning of primitive units" thus forming a crucial initial stage for language acquisition. More specifically in relation to music, Trehub (2001) has also demonstrated that melodic contour is a crucial element for infant perception of melodies.

> Melodic contour is an especially durable feature of melodies. Remarkably, infants can identify changes in contour when the standard and comparison melodies are separated by a full 15 seconds or when distractor tones are inserted between standard and comparison sequences. In fact, pitch contour seems to be the most salient musical feature for infant listeners. It may also be the most salient feature of mothers' speech to prelinguistic infants. Evidence from adult listeners is consistent with contour processing as fundamental and relatively impervious to musical experience. During contour-processing tasks, for example, the amplitude and latency of event-related potentials do not differ for musicians and nonmusicians; during interval-processing tasks, however, experience-dependent differences are apparent.[8]

Trehub's finding that "pitch contour seems to be the most salient musical feature for infant listeners" shows that it plays a foundational role in our musical development. This is consistent with evidence from adult melody perception. Judy Edworthy (1985) found that melody (intonation) has greater immediate perceptual salience for listeners than

---

[7] Caroline Palmer and Sean Hutchins, "What is musical prosody?" In *The Psychology of Learning and Motivation, Vol. 46: Advances in Research and Theory*, ed. Barry Ross (San Diego, CA, Academic Press, 2006): 245-278.
[8] Trehub, "Musical predispositions in infancy," 4.

elements of intervallic structure, and that it may well prove to be the first step in the encoding of a pitch pattern.[9] I will expand on these points in Chapter 3, and in the analyses of pitch contour and intonation in Chapter 4, where I show that the global attributes of intonation are just as important to our perception and cognition of melody as are local pitch events. The important point here is that IDS exhibits many prosodic features at a formative and possibly universal phase in human development, prior to the sorts of differentiation that take place across cultures, and where the tonal features of (infant) song and speech are closely intertwined.

The features of IDS are species-specific and suggest an early origin in our hominin lineage. There is still considerable debate around these issues, but most scholars point to the *social* dimensions of these communicative channels and the affordances they may have had in culture-rich environments. Dean Falk claims that the continuous, positively valenced, affective vocalizations that characterize mother-infant interactions in modern humans are absent in our nearest extant relatives, chimps and bonobos, and suggests that these features arose in part to 'substitute' for the mother's proximity as hominin infants became more altricial and less able to cling to the mother.[10] Similarly, Ellen Dissanayake claims that the use of 'motherese' in affect regulation and co-regulation in mother-infant interactions provides a rationale for music's roles in "social regulation and emotional

---

[9] Judy Edworthy, "Interval and contour in melody processing," *Music Perception* 2, no. 3 (Spring, 1985): 375-388.

[10] Altricial refers to the extended period of dependency that characterizes mother-infant relationships in humans and other primates. See: Dean Falk, *Finding our tongues: mothers, infants and the origins of language* (New York, Basic Books, 2009).

conjoinment."[11] The sounds and songs that infants learn through childhood are important to their subsequent perception and interpretation of their sound worlds. And they learn these sounds through imitation of their parents and care-givers. The sorts of categories they learn and how they learn them is critical for understanding the role of prosodic pitch features and structure later on. For there is no doubt that "[o]ur native sound system leaves an imprint on our minds. That is, learning a sound system leads to a mental framework of sound categories for our native language or music. This framework helps us extract distinctive units from physical signals rich in acoustic variation. Such frameworks are highly adaptive in our native sonic milieu."[12] The social dimension to music acquisition and development underscores the need for usage-based models. To understand how the categories of music and language are learned—in other words, the role of cultural inheritance—we need to study specific cultures in context, and using a comparative method that pinpoints similarities across cultures.

Another important point about IDS is that it is produced by mothers specially *for* infants as a tool to aid cognitive development and the regulation of emotions: "From infancy through the preschool years, a mother's verbal interactions with her child seem to have didactic as well as social-regulatory goals."[13] And the role of song and speech to infants

---

[11] Ellen Dissanayake, "Antecedents of the Temporal Arts in Early Mother-Infant Interaction," in *The Origins of Music*, ed. Nils L. Wallin, Björn Merker, and Steven Brown (Cambridge, MA: MIT Press, 2000) 463-472.
[12] Aniruddh Patel, *Music, language and the brain* (New York, Oxford, 2008), 9.
[13] Sandra Trehub, "Communication, Music, and Language in Infancy,". In *Language, Music and the Brain: A Mysterious Relationship*, ed. Michael Arbib (Cambridge, MA: MIT Press, 2013), 463-479.

is very similar. It is crucial that infants engage and learn from these experiences if they are to develop normally:

> Culturally typical intonation patterns (i.e., speech melodies) are distinguishable in the babbling of preverbal infants from different language backgrounds (de Boysson-Bardies et al. 1984). The vocalizations of infants from English-speaking environments are dominated by falling contours, and those from French-speaking environments are dominated by rising pitch contours (Whalen et al. 1991). Remarkably, three-month-olds produce spontaneous imitations of prosodic contours during mother-infant interactions (Gratier and Devouche 2011). Even more remarkably, newborn cries are influenced by the ambient language *in utero*, with French newborns producing more rising contours and German newborns producing more falling contours (Mampe et al. 2009). These studies seem to suggest that preverbal infants accord priority to global or suprasegmental aspects of speech despite their ability to differentiate segmental aspects.[14]

All of this evidence points to the fact that the learning of pitch features in song and speech happens very early, perhaps even in utero, and continues unabated through childhood. Infants quickly develop culture-specific propensities that are not evident at first but which are learned through their mothers' (and later, other caregivers) songs. For instance, Trehub points out that, "there have been no demonstrations of infant listening preferences for melodic intervals (i.e., combinations of successive tones) that Western adults consider consonant or dissonant."[15] This finding shows that pitch categories and their organization and structure are not universal, but are culture-specific. Another

[14] Trehub, "Communication, Music, and Language in Infancy," 472.
[15] Trehub, "Communication, Music, and Language in Infancy," 466.

important point is that infant learning is multi-modal and takes place through gestural, visual, and auditory stimuli. This suggests that domain-general capacities are at work in early childhood development and that an element of statistical social learning buttresses these capacities. A further element discussed by Trehub is that "infants accord priority to global or suprasegmental aspects of speech." This important point will be discussed in Chapters 3 and 4 where I show that such global features remain important in the perception of adult melodies. Usage-based studies show that we learn speech elements based on function and that utterances are usually perceived and encoded holistically by infants. The same may well be true of music. This process of 'chunking' information appears to be domain-general and is a property of both vision and audition. These holistic and Gestalt-like processes require us to rethink our conception of music as a *discrete* combinatorial system. It may well be combinatorial, but the elements combined may differ by degree and kind.

In sum, this research on IDS shows that song and speech share important prosodic features, that these features are acquired early on in ontogeny, that pitch categories may consist of categories that are holistic gestalts, and that these are *learned* and culture-specific. These findings are consistent with a usage-based approach that focuses on the role of domain-general, social-cognitive learning mechanisms. The IDS research shows that there is considerable overlap between the domains of music and language in infancy, and this evidence has been corroborated by studies in neuroscience, some of which I review in 1.3 below.

1.3 Neuroscience

Neuroscientific evidence shows that there is significant neural overlap for the processing of instrumental music and language. Interpreting this evidence, Aniruddh Patel (2003) argues that language and music rely on the same mechanisms, even in cases where they recruit different areas of cortex.[16] Isabelle Peretz and Robert Zatorre (2005), on the other hand, contest this idea, arguing that these mechanisms are insufficient to explain the complexity and distinctiveness of music processing.[17] Peretz states that, "musical abilities might well be distinct from language and be subserved by specialized neural networks under the guidance of innate mechanisms. Accordingly, any given individual would be born with the potential to become musically proficient."[18] At present we have limited means with which to resolve this disagreement, at least in neuroscience itself. Current methods are limited in their ability to investigate the complexity and distinctiveness of these behaviors. For instance, it seems impossible to localize the *functions* of the underlying mechanisms, suggesting that localization itself gets us little closer to a cognitive explanation. The case is especially difficult for song and speech where processing occurs via the same signal (melody/intonation).[19] Many experimenters have tried to control for this by comparing instrumental music with language, rather than song

---

[16] Aniruddh Patel, "A new approach to the cognitive neuroscience of melody," in *The Cognitive Neuroscience of Music*, ed. Isabelle Peretz and Robert Zatorre (New York: Oxford, 2003), 326-345.

[17] Isabelle Peretz and Robert J. Zatorre, "Brain organization for music processing," *Annual Review of Psychology*, 56 (February 2005): 89-114.

[18] Isabelle Peretz, "The biological foundations of music: Insights from congenital amusia," in *The Psychology of Music*, Third Edition, ed. Diana Deutsch (San Diego, CA: Academic Press, 2012): 551-564.

[19] "[N]o one study has yet directly compared spoken language, sung language, and vocalizations within the same design and with the same participants." Daniele Schön, Reyna Leigh Gordon, and Mireille Besson, "Musical and Linguistic Processing in Song Perception," The Neurosciences and Music II: From Perception to Performance, *Annals of the New York Academy of Sciences* 1060 (December 2005): 75.

with speech.[20] The result is that the relationship between the latter remains still requires investigation. Even in cases where instrumental music and language are shown to recruit different areas of cortex, there is no guarantee that different mechanisms are involved. Far more research is required—and not just in brain localization—in order to establish what these mechanisms are.

A crucial theoretical issue that studies in neuroscience have shed much light on is that of modularity. [21] Nativists have long advanced the position that music and language occupy distinct 'modules' in the brain. In the 1980s and 90s there was optimism that brain localization studies would reveal specialized and distinct modules that were genetically predisposed.[22] This idea was undermined by a lack of supporting empirical evidence and

---

[20] In *Music, Language and the Brain* (2008), Patel compares instrumental music and language rather than song and speech, precisely because of the more obvious contrasts between the former (and with respect to pitch patterning in particular).

[21] Many in cognitive neuroscience assume there is a biological basis to cognitive capacities shared by all healthy humans and that such capacities are universal. Biological processing is taken to be definitive of specificity for either language or music and this measure is usually given precedence over 'behavior' (in part because it escapes neat classification). Cognitive psychologists label specificity through recourse to experiments that display a 'double dissociation' between two behaviors. Here it is important to distinguish between domain-specificity (as exhibited by such double dissociations) and 'modularity.' "A domain-specific operation is a distinct mechanism that deals with a particular aspect of the input and does this either exclusively or more effectively than any other mechanism. What individuates a module is its functional specialization" (Isabelle Peretz, "Music, Language and Modularity Framed in Action," *Psychologica Belgica* 49, 2-3 (2009): 159-160). Domain-specific operations need not imply music specificity and may not involve special purpose learning mechanisms. "Domain-specificity may either emerge from general learning processes or result from the nature of the input code" (Ibid). Studying the specificity of song from this perspective means attending to each component part of the system rather than the faculty as a whole. This allows neuroscientists to sidestep the problem of the ontological unity of music and difficulties of global cognition. The component we consider here is pitch, and the rationale extends Darwin's insight—substantiated now by many studies of IDS and prosody—that song and speech share crucial design features (Tecumseh Fitch, "The biology and evolution of music: A comparative perspective," *Cognition* 100 (2006): 173-215).

[22] Jerry Fodor's *Modularity of Mind* (Cambridge, MA: MIT Press, 1983) proved to be a very influential text for computationists because it provided significant theoretical support for the nativist position in language and music. Studies by Lerdahl and Jackendoff (1983) and Narmour (1990) both postulated a role for modules in their theories. Weiskopf summarizes Fodor's initial conception of modularity as follows:

accumulated information to the contrary that showed music and language to be the product of processing on a global scale. Fiona Cowie reviews the evidence for localization as it relates to nativism as follows:

> In sum, the neuroscientific evidence currently available provides no support for linguistic nativism. The suggestion that localization of function is indicative of a substantial degree of innate prespecification is no longer tenable: localization can arise in many different ways. In addition, linguistic functions do not seem to be particularly localized: language use and understanding are complex tasks, involving many different brain areas – areas that are in at least some cases implicated also in other tasks. It is hard to see how to reconcile these facts with the Chomskyan postulation of a monolithic 'language organ', the development or 'growth' of which is controlled largely by the genes. Finally, the fact that complex functions can be learned and carried out by areas of brain that are innately 'prewired' (if at all) to do quite different sorts of processing indicates that such competences can be and are acquired without any inborn, task-specific guidance. This is not, of course, to say that language is one of the competences that are acquired in this way. For all the current evidence shows, many areas of cortex in which language develops may indeed be 'prewired' for that task: linguistic nativism is still consistent with what is now known. It is, however, to suggest that although there may be other reasons to be a linguistic

---

"According to MOM, modules are domain-specific, mandatory in their operation, fast, and informationally encapsulated. They have relatively shallow or impoverished outputs, and extramodular processes, especially those involved in belief formation, have little or no access to the intermediate representations they use in computing those outputs. Their ontogeny exhibits a characteristic pacing (e.g. there are critical periods for the development of language and normal vision), and they have regular patterns of breakdown or malfunction. Finally, they are associated with a relatively stable neural architecture, which explains this maturational pacing and pattern of deficits, since different parts of the brain may have different rates of growth and development, and different propensities to recover from injury. The neural architecture associated with a module also explains its informational encapsulation, since the absence of an anatomical pathway between two parts of the brain prevents either one from accessing the other's informational contents." Daniel A. Weiskopf, "A critical review of Jerry A. Fodor's *The mind doesn't work that way*," *Philosophical Psychology* 15, No. 4 (2002): 555.

> nativist, general considerations to do with brain organization or development as currently understood give no especially support to that position.[23]

The fact that 'localization may arise in many different ways' means that we should be cautious in assuming a correlation between localization and specificity. The wide and complex distribution of brain activity undermines the notion of discrete modules, and it supports the idea of domain-general capacities as the underlying mechanisms that drive higher cognitive functioning, including both music and language. Still, even if modularity now seems implausible, this does not negate the possibility of music- or language-specificity entirely. For instance, recent evidence has emerged suggesting that pitch plays an important and possibly special role in music.

The best evidence for music specificity comes from studies of *congenital amusia*. This refers to a "lifelong deficit in melody perception and production that cannot be explained by hearing loss, brain damage, intellectual deficiencies, or lack of music exposure."[24] The crucial dimension to this deficit, found it roughly 2-3% of the population, is *tonal encoding of pitch*.[25]

> Tonal encoding of pitch is the prime candidate [for congenital amusia] as it appears unique to music. Pitch variations generate a determinate scale in music but not in speech intonation

[23] Fiona Cowie, "Innateness and Language," in *The Stanford Encyclopedia of Philosophy,* Summer 2010 Edition, ed. Edward N. Zalta (Stanford, CA: Stanford University), http://plato.stanford.edu/archives/sum2010/entries/innateness-language/ (accessed June 3, 2013).
[24] Peretz, "The biological foundations of music: Insights from congenital amusia," 552.
[25] See Peretz, "The biological foundations of music: Insights from congenital amusia." These statistics are based on tests using the Montreal Battery of Evaluation of Amusia (MBEA) discussed in this article.

> contours. Furthermore, the use of fixed and discrete pitches that are mapped onto musical scales are nearly universal (Dowling & Harwood 1986). The pitch sets used in a given musical culture remain intact from generation to generation, even in the absence of acoustic instruments or notation. The vocal play of 6- to 12-month-olds that leads to singing is distinguishable from the vocal play associated with incipient speech, both in its use of stable pitch levels on vowels and in its rhythmic organization in terms of a regular beat pattern (Dowling, 1984, 1999, p. 611; Dowling and Harwood, 1986, p. 147). This finite pitch set enables the generation of an infinite number of musical structures.[26]

The idea that tonal encoding of pitch applies primarily to discrete pitch categories is not the whole story. There is some evidence that even fine-grained properties of intonation contours are affected.

> In both French and English, intonation is used to convey a question or a statement. Amusics have little difficulty in distinguishing these (e.g., Hutchins, Zarate, Zatorre, & Peretz, 2010; Patel, Wong, Foxton, Lochy, & Peretz, 2008) although they may show mild impairments when these pitch changes are subtle (Hutchins, et al., 2010; Liu, Patel, Fourcin, & Stewart, 2010) or require memory (Patel et al., 2008). Similarly, amusics may experience difficulties when comparing lexical tones taken from Mandarin or Thai words (Tillmann, et al., 2011). […] Thus, amusics may show a deficit in processing pitch information in speech, but this deficit is generally mild.[27]

---

[26] Peretz, "The biological foundations of music: Insights from congenital amusia," 555-556. See also: W. Jay Dowling and Dale L. Harwood, *Music cognition* (New York, NY: Academic Press, 1986); W. Jay Dowling, "Development of musical schemata in children's spontaneous singing," in *Cognitive processes in the perception of art*, ed. W.R. Crozier and A.J. Chapman (Amsterdam, The Netherlands: North-Holland, 1984): 145-163; W. Jay Dowling, "The development of music perception and cognition," in *The Psychology of Music*, Second Edition, ed. Diana Deutsch (San Diego, CA: Academic Press, 1999): 603-625).

[27] Peretz, "The biological foundations of music: Insights from congenital amusia," 554. See: Sean Hutchins, J.M. Zarate, Robert J. Zatorre, and Isabelle Peretz, "An acoustical study of vocal pitch matching in

Peretz concludes her review amusia studies that tonal encoding of pitch is in all likelihood music-specific. And yet she also provides important evidence for related pitch-processing deficits amongst speakers of tone languages.[28] For instance, "speakers of tone languages such as Mandarin may experience musical pitch disorder despite early exposure to speech-relevant pitch contrasts […]. Our results with amusic speakers of Mandarin suggest that perhaps the same genes are involved in both tone language and musical pitch processing."[29] This is a fascinating finding because it shows how complex pitch ordering may not be entirely music-specific, and that the prosodic features of tone languages seem to share similar properties to musical melodies.

Tonal encoding of pitch is vital, it seems, to the perception and cognition of categorical properties of discrete pitch, or what I will later term 'pitch targets' (see Chapter 3). But there are others features of pitch patterning in song that may also be affected by these sorts of deficits. These 'gradient' elements have not yet been investigated because of the over-emphasis, in my view, on discrete pitches and intervallic structures in common practice tonal and instrumental music (i.e. on notational conventions). Most cognitive scientists have designed experiments on the assumption that 'melody' is equivalent to a

---

congenital amusia," *Journal of the Acoustical Society of America*, 127, No. 1 (2012): 504-512; Aniruddh D. Patel, Meredith Wong, Jessica Foxton, Aliette Lochy, and Isabelle Peretz, "Speech intonation perception deficits in musical tone deafness (congenital amusia)," *Music Perception* 25, No. 4 (2008): 357-368; Fang Liu, Aniruddh D. Patel, Adrian Fourcin, and Lauren Stewart, "Intonation processing in congenital amusia: discrimination, identification and imitation," *Brain* 133, No. 6 (April 2010): 1682-1693.

[28] Barbara Tillmann B, Dennis Burnham, Sebastien Nguyen, Nicolas Grimault, Natalie Gosselin, and Isabelle Peretz, "Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages," *Frontiers in Psychology*, 2 (June 2011): 1-15; Yun Nan, Yanan Sun, and Isabelle Peretz, "Congenital amusia in speakers of a tone language: association with lexical tone agnosia," *Brain* 133, No. 9 (August 2010), 2635-2642.

[29] Peretz, "The biological foundations of music: Insights from congenital amusia," 560-561.

series of discrete pitches patterned in time.[30] Where contour is factored into such studies it is usually taken to be the 'shape' that this series of pitches gives rise to. But transposing (sung) melodic contours to instrumental (usually keyboard) music ignores the gradient features that are crucial to perceiving affective and paralinguistic meanings, not to mention the meanings conveyed by stress and intonation. This bias is evident in many studies. For instance, in their comparative study of French and English melody and rhythm, Patel, Iversen, and Rosenberg (2006) used *prosogram* to standardize pitch contours into level tones "for maximum comparability to music."[31] It was assumed that continuously unfolding contours are more language-like. But should pitch patterning in instrumental (keyboard) music be considered the standard for understanding this capacity? Are song melodies not equally continuous in their unfolding?

There are several problems with these sorts of approaches modeled on instrumental music. We need to recognize that pitch patterning in song (i.e. vocal music) tends to be more gradient than for many instruments, especially the keyboard instruments used in so many laboratory experiments. There are at least four reasons for considering instrumental music an inappropriate model for measuring the relationship of song (music in its embodied form) to language:

[30] Schön et al., "Musical and Linguistic Processing in Song Perception."

[31] Aniruddh D. Patel, John R. Iversen, and Jason C. Rosenberg, "Comparing the rhythm and melody of speech and music: The case of British English and French," *Journal of the Acoustical Society of America* 119 (2006): 3039.

(1) Song melodies tend to have glides and contours that are not discrete, and these gradient features may be categorical in and of themselves.[32]

(2) Song melodies do not always conform to the intervallic structures specified in musical scores or realized on instruments tuned to specific standard pitches (such standardization is of course very recent in the West).

(3) Singers tend to use micro-prosodic pitch elements for expressive purposes that are not factored into models of discrete pitch. These include intonational and 'paralinguistic' features that are excluded from generative models.

(4) Pitch patterning in song is determined in part by 'linguistic' pitch factors. These may include elements of speech tone, intonation and pitch-accent, depending on the language sung.[33]

None of these factors are relevant to instrumental music so they are generally ignored in experimental designs, and yet they are vital to our study of song and the capacity for music as a feature of human cognitive abilities. We have hardly begun to investigate the non-discrete elements of pitch patterning in song. This is in part the rationale for treating melody *as* prosody in this dissertation. In Chapters 3 and 4 I discuss considerable evidence that song and speech utterances have gradient features that are processed

[32] The term 'discrete' is most often used to refer to level pitch categories, but such categories are not present in many songs and so I introduce the concept of 'pitch target' in Chapter 3.

[33] I consistently use the descriptor 'linguistic' for elements of speech tone, intonation and pitch accent. This is the convention and it is useful here because it points to features of song that have largely been of peripheral interest to music theorists. But it is debatable whether or not prosodic elements are inherently 'linguistic' rather than 'musical.'

categorically and as larger chunks of sound.[34] The usage-based framework I propose treats song as a *complex adaptive system* that is fit for purpose. This is in stark contrast to standard computationist models that predicated on mentalistic grammars of discrete, time-independent symbols; gradient pitch categories cannot be accommodated into such systems.

To conclude, the processing of speech and song occupy similar brain regions but with different degrees of activation in these regions. Instrumental music may rely on an order of processing different from song, and so it is as yet unclear to what extent vocalization plays a role in localization. Complementary methods of investigation, including those developed in this dissertation, are likely to provide a more nuanced picture of these relationships. In the next section I discuss prosody as an important component of evolutionary studies. These studies also provide a context for debates on specificity and the biological basis of song as the product of multiple interrelated capacities.

1.4 Evolution

There has been a resurgence of interest in evolutionary studies of music over the past two decades. Much of this work has focused on various scenarios for the origins of music and language, and some have uncovered a deep history. It was Charles Darwin (1871) who made a strong case for prosody as the crucial link between speech and song.

[34] There are nevertheless some differences between pitch patterning in song and speech that must also be accounted for. One of these, it has been found, is the degree of pitch discrimination. In speech large pitch intervals are the norm, while in song we find a combination of both large and small pitch intervals. This is described by the notion of pitch adjacency, whereby intervals are instantiated in tonal pitch space.

> As we have every reason to suppose that articulate speech is one of the latest, as it certainly is the highest, of the arts acquired by man, and as the instinctive power of producing musical notes and rhythms is developed low down in the animal series, it would be altogether opposed to the principle of evolution, if we were to admit that man's musical capacity has been developed from the tones used in impassioned speech. We must suppose that the rhythms and cadences of oratory are derived from previously developed musical powers. We can thus understand how it its that music, dancing, song, and poetry are such very ancient arts. We may go even further than this, and, as remarked in a former chapter, believe that musical sounds afforded one of the bases for the development of language.[35]

One hundred and forty years have passed since the publication of Darwin's *Descent of Man* (1871), and yet his theory of the origins and evolution of music and language remains largely conjectural. Why is this so? What factors have heretofore confounded our efforts to disentangle speech and song, language and music, and how might we untie this evolutionary knot? Darwin made several observations that assist us in this investigation. First, his suggestion that we study song as the basis for instrumental music—rather than the other way round—should be taken seriously. Second, he suggested that tonal prosodic features are crucial to both music and language and may well have been the basis for the development of both. These observations resonate with the focus of this study and with evidence from IDS and neuroscience already reviewed in this chapter. But Darwin's theory of evolution by natural selection also provides a context for understanding music as a 'complex adaptive system,' the very framework that will be developed in later chapters. Song constitutes a specialized system that is neither wholly separable from nor

[35] Charles Darwin, *The Descent of Man* (1871, Repr. New York: Penguin, 2004): 638-9.

wholly dependent upon other capacities and natural selection is the only mechanism that produces complex adaptation of this sort.[36] A 'complex adaptive system' is one composed of multiple interacting parts where the details of the parts' structures and arrangements suggest design to fulfill a specific function. There is an extremely low-probability of complex abilities such as these having a non-selectionist explanation. If music is an adaptation, then it must have arisen by these processes either in whole or in part. Understanding song as a 'complex adaptive system' places restrictions on how we theorize music and mind. In particular, it suggests that simplistic single-origin models for the evolution of music are implausible.[37] These are the sorts of models to which nativists are bound. Usage-based theorists like Michael Tomasello have instead proposed a gradual, incremental accrual of the capacities we now know to comprise music.[38]

It is worth stressing the point that historical processes are too often ignored in theories of music's origins and evolution, with the result that proximate and ultimate causes are conflated. The evolutionary psychology paradigm in music cognition studies has tended to ignore the complications of historical processes of change and interaction in favor of differentiated Fodorian modules that evolved in unspecified 'environments of

[36] For a classic discussion of adaptive complexity and language see: Steven Pinker and Paul Bloom, "Natural Language and Natural Selection," *Behavioral and Brain Sciences* 13 (1990): 707-784.
[37] For further discussion see: Thomas M. Pooley, "Dismantling music: Reductionist models and evolutionary explanations in music cognition," in *Topicality of Musical Universals/Actualité des Universaux Musicaux*, ed. Jean-Luc Leroy (Paris: Éditions des archives contemporaines, 2013): 43-48.
[38] See: Tomasello, *Constructing a Language*; Michael Tomasello, *Origins of Human Communication* (Cambridge, MA: MIT Press, 2010).

evolutionary adaptedness.'[39] The atomistic and reductionist models of experimental psychology and cognitive neuroscience favor such models because they are far easier to 'test' and because of the continued emphasis on nativist explanations.

Evolutionary psychologists have proposed a 'music module' and a variety of single-origin hypotheses to explain its evolution. These speculative theories share the flawed assumption that music's 'design features' underwent independent evolutionary histories and that the present condition and relations of these features have remained constant over the course of human history. The relatively diffuse definitions of music in current use confuse the matter further, and the absence of theoretical engagement with the problem of specificity only compounds this. Tecumseh Fitch reminds us of the importance of an historical understanding of the 'music faculty.'

> Although all of the mechanisms involved in music perception and production may be grouped together, for convenience, as 'the music faculty' or 'the capacity for music', it is important to remember that different components of this capacity may have different evolutionary histories. Thus, discussing 'Music' as an undifferentiated whole, or as a unitary cognitive 'module', risks overlooking the fact that music integrates a wide variety of domains (cognitive, emotional, perceptual, motor…) may serve a variety of functions (mother-infant bonding, mate choice, group cohesion…) and may share key components with other systems like language or speech. Thus,

[39] See: Jerome Barkow, Leda Cosmides, and John Tooby, *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (New York: Oxford, 1992).

> questions like 'When did music evolve?' or 'What is music for?' seem unlikely to have simple unitary answers.[40]

A key difficulty in accounting for any complex capacity is describing its phylogenetic history. How can we make sense of several intertwined capacities with co/in-dependent histories? Trade-offs and interdependencies are characteristic of all human biological functions, a fact that complicates the whole notion of adaptation. But the question of integration should not be dismissed a priori. Rather, and if we are to take a phylogenetic explanation seriously, we need to think about the constraints imposed by these multiple components on one another and the constraints imposed by selection on the organism as a whole. In order to dismantle music into its design features, capacities, or traits we must take into account their co-evolution over time, and this means understanding their functional interdependence as well as their independent existence. Reductionist methods in music cognition run the risk of inappropriate atomization. To move beyond 'just-so stories' of music's origins and evolution—predicated as they are on a monolithic and presentist notion of music—we need rigorous and empirically grounded criteria for studying the music faculty across disciplines. To reify individual traits as *metonymic* of the music faculty as a whole is to reduce an ancient, complex and biologically encoded behavior to a simple formula. The challenge is rather to establish a framework for studying music as a complex yet irreducible behavior that is the product of multiple interacting forces and selective pressures through phylogeny. One way of achieving this is by focusing not on the nature of the capacities as isolated entities, but on their

[40] Fitch, "The biology and evolution of music: A comparative perspective," 173.

integration. To do this we must recognize the extent to which sounds are embedded in social contexts rather than biologically determined by modules. This is the very purpose of a biocultural musicology: to recognize both the inheritance and the emergence of behaviors through ontogeny and phylogeny.

This dissertation proposes a framework for understanding how one aspect of the complex system that is song—pitch—is used and is structured to meet its multifarious communicative functions. Having already reviewed the biological evidence for pitch patterning in song and its evolutionary basis, we must now establish the nature of musical diversity across cultures. Nativist and especially generativist accounts of music and language do not factor in the extraordinary diversity we find in world cultures. But we need to understand this diversity and how the social processes that give rise to it enable the acquisition of music and language. To do so we need to study *direct* evidence of musical behaviors rather than artifacts of Western culture. This means adopting a biocultural perspective that explains, paradoxically, both the nature of musical diversity and the status of its universals.

PART 2: THE NATURE OF MUSICAL DIVERSITY

1.5 The need for cross-cultural studies

The diversity of musical expression can only be explained by virtue of its cultural origins. If there were inflexible biological constraints on the nature of music then we would not see the diversity and indeed the creativity so indicative of music making. Understanding the role of culture is therefore a priority and this can only be accomplished through detailed studies of individual cultural manifestations of music, and through comparison of such manifestations across cultures. Furthermore, universal features of human cognition cannot be understood through an ethnocentric lens. For too long now, scholars of music cognition have used Western art music and its systems of notation and analysis to model music and mind. Idiosyncratic features of Western tonal and metrical systems have been taken to be normative of music and therefore universal while other parameters, such as timbre, have been ignored.

Compositions by Bach, Haydn, Mozart, Beethoven, Chopin, Schumann, Brahms, and many other eighteenth and nineteenth century composers were the basis for analytical work in early cognitive music theory. Even today, it is to the repertoire of popular and classic tonal music that experimenters turn when modeling music and mind. Not only are the procedures of common practice tonality all too frequently the focus of investigation, but they are assumed by many investigators to provide evidence of 'universal' categories appropriate to the study of non-Western musics too. We see this bias in the use of staff notation for the transcription and analysis of world musics, and in the assumptions that music systems consist of time-independent objects or works. To make headway on important questions in cognitive science—including topics of universals, evolution, diversity, and shared capacities—we must broaden the scope of our investigations to include new repertories and new methods—the two in tandem—and we must move to studies of performance and process rather than the so-called ‘competence’ objectified in musical scores.

One consequence of taking common practice tonality as normative of music is that non-western cultures have received only sporadic study, and so a genuine cross-cultural perspective on music and mind remains elusive. In their review of scholarship in comparative music psychology, Aniruddh Patel and Steven Demorest (2013) describe the overwhelming emphasis on Western music as follows:

> Music psychology has, until recently, largely ignored [musical] diversity and focused almost entirely on Western music. This was a natural tendency given that most of the researchers in the field were enculturated into Western musical styles. Unfortunately, theories and research findings based solely on a single culture’s music are severely limited in their ability to tell us about music cognition as a global human attribute. This is why comparative approaches to music psychology, although relatively new, are critical to our understanding of music cognition.[41]

Catherine Stevens (2012) has also drawn attention to the importance of studying world musics rather than focus on Western styles.

> World musics provide a laboratory for the investigation of music perception and cognition where there has been little or no prior exposure to a particular musical system. Such cross-cultural research is important for at least three reasons. First, as the West encroaches, there is an

[41] Aniruddh Patel and Stephen Demorest, “Comparative Music Cognition: Cross-Species and Cross-Cultural Studies,” in *The Psychology of Music*, Third Edition, ed. Diana Deutsch (London: Academic Press, 2012): 647-682.

> irreversible loss of cultural diversity (Huron, 2008) and with it knowledge of different musical systems and psychological processes. Second, because theories of perception and cognition have been conceived in Western settings and within a monoculture, psychological theories need to be evaluated and challenged in a variety of contexts. Third, the bias to the West has inhibited the take up of theory and empirical findings that arise from the way music is created, performed, perceived, used, and learned in diverse cultural settings. An extreme example of this has been a concentration during the 20th century on Western tonal art music that is performed for a seemingly passive audience or listened to via audio recording. Such a narrow view would not have occurred if we had considered situations where music, movement, dance, and visual art are inseparable and music is social, dynamical, and interpersonal.[42]

These two review articles point to the need for more systematic investigation of music in world cultures that aims not only at discovering culture-specific knowledge but also cross-cultural commonalities and differences. Without such studies we run the risk of ethnocentrism. For as David Huron points out: “It may be that all of the important lessons regarding music can be found in Western music. But who would be so presumptuous as to assume this to be the case before we investigate the matter thoroughly?”[43]

Music cognition is not currently well placed to offer empirical support for theories of universals and evolution. This bias is in part the rationale for the case study of Zulu song prosody presented in Chapter 4. This case study counters not only the emphasis on common practice tonality, but also the use of analytical procedures reliant on staff

---

[42] Catherine Stevens, “Music Perception and Cognition: A review of recent cross-cultural research,” *Topics in Cognitive Science*, 4 (2012): 653-654.
[43] David Huron, *Sweet Anticipation* (Cambridge, MA: MIT Press, 2006): 379.

notation. By using spectrographic methods of analysis, as well as the 'neutral' analytical categories devised by phoneticians and phonologists, I aim to move beyond the biases inherent in many theories music cognition. I build on the work of early comparative musicologists and ethnomusicologists who focused on similar sorts of problems (see Chapter 5). However, if we take ethnography to be a defining feature of much recent ethnomusicology, then there are good reasons to conceive this project as a study in music theory. Indeed, there is no reason why studies of world music should not stake a central claim to acceptance in this discipline today.

The focus of comparative studies tends to be on questions of similarities or universals. For some time the question of universals was thought to include absolute synchronic universals that could—with sufficient descriptive accuracy—be identified in all musics. This seems to have been a misguided enterprise and so more recent studies have focused on underlying mechanisms instead. Having already reviewed the evidence for such mechanisms in the first part of this chapter, I now focus on two approaches to song and speech that characterize recent work in cognitive science. These fall under two main categories: computationism and usage-based theory.

1.6 Computationism

That song and speech are universal attributes of human communication is uncontested. What is less clear, however, and what perplexed Darwin, is why? In *The Descent of Man* he writes that human musical abilities are "amongst the most mysterious with which

[man] is endowed."[44] This is not so of speech. Spoken utterances are the product of a communicative system that conveys both propositional and affective information. Song relies on similar mechanisms but its messages are often redundant, conveying the sort of propositional information we expect of language, but in different contexts and with different results. What purposes might song serve, and where should we look to find its most salient features?

The Chomskian revolution in cognitive science shifted our focus from surface structures to 'deep structures' and an underlying 'mental grammar.' The insight that the structure of cognition is complexly ordered according to principles not immediately apparent in the acoustic signal is a vital one for all studies of music, language and mind because it reminds us that descriptive and classificatory studies will get us no closer to understanding human cognition, however exhaustive they might be. Nevertheless, Chomsky's rationalist approach—with its principles of nativism, mentalism and a universal grammar—do not accord with a great deal of empirical evidence now available in the cognitive sciences. There is one basic problem with rationalist models that I will return to throughout this dissertation: they do not admit of the gradience and variation we find across cultures. This is because they are predicated on a combinatorial system that computes a discrete set of mental symbols that must be shared by all humans. Computationism relies on a formal symbol logic derived from the mathematical reasoning of Alan Turing. He had the influential idea that mental processes are

[44] Ibid.

computations that are syntactically driven. A computation is "a formal operation on syntactically structured representations. Accordingly, a mental process, qua computation, is a formal operation on syntactically structured mental representations."[45] Chomsky's generative grammar relies on this computational framework. It is also Cartesian in the sense that humans possess a creative component that enables them to generate a potentially infinite number of combinations from a limited number of symbols and to structure according to a limited number of rules or constraints that are universal across cultures. This 'discrete combinatoriality' is predicated on the notion that both pitch and rhythm are perceived as discrete mental symbols that may be manipulated independently of social contexts to produce meaningful utterances. There are many reasons to be skeptical of computationism. Apart from the fact that its nativist assumptions are largely unsupported by studies of language acquisition and brain localization, the cognitive reality of mentalistic grammar has also yet to achieve empirical support. For our purposes there is another important lacuna: prosodic features are seldom discrete. Pitch patterning is generally continuous and gradient, and displays considerable variation across cultures. We therefore need an alternative to computationism that takes account of these features: in this dissertation I propose a usage-based theory of music.

1.7 Usage-based theory

Michael Tomasello and Joan Bybee are two prominent advocates of usage-based linguistics and they have argued that domain-general learning abilities are the driver of

[45] Jerry Fodor, *The mind doesn't work that way: The scope and limits of computational psychology*. (Cambridge, MA: MIT Press, 2001): 11.

language acquisition and use.[46] Tomasello has countered many of the assumptions of generative and nativist accounts of language acquisition in his book *Constructing a Language* (2003). He describes two main social-cognitive skill sets that enable children to acquire language without a 'language acquisition device.' These include: *intention-reading* (theory of mind) and *pattern-finding* (categorization). Similarly, Bybee (2010) discusses categorization, chunking, rich memory storage, analogy and cross-modal association—all of which may be subsumed within Tomasello's two broad skill sets. These approaches stress that language emerges from use rather than as the product of an innate grammar.

The usage-based approach is consistent with the evidence presented in this chapter since no assumptions are made about innate capacities or the cognitive reality of mental representations. I will show how a similar approach may be applied to music. Fortunately there are several music theorists who have investigated similar domain-general learning mechanisms. The first, perhaps, was Leonard Meyer who argued that, "one cannot comprehend and explain the variability of human cultures unless one has some sense of the constancies involved in their shaping."[47] Some of these constancies, as we will review them in Chapter 3, are Gestalt principles of similarity, proximity, and common fate; the very same principles that Eugene Narmour was to develop in his implication-realization

---

[46] Tomasello (2003) presents a 'global' theory of usage-based linguistics whereas Bybee (2010) discusses select mechanisms of special interest to her. See: Joan Bybee, *Language usage and cognition* (Cambridge: Cambridge University Press, 2010).

[47] Leonard Meyer, "A universe of universals," *Journal of Musicology* 16, No. 1 (1998): 3. These constancies included, for instance, the Gestalt grouping laws of similarity, proximity, and common fate that I will discuss in much more detail in Chapter 3.

model.[48] Further evidence comes from studies of statistical learning, many of which are outlined by David Huron (2006) in *Sweet Anticipation*, and from the acquisition studies reviewed in the first part of this chapter. Studying these mechanisms gets us away from simplistic notions of absolute synchronic universals, as Meyer was careful to point out:

> To explain why human beings in some actual cultural-historical context think, respond, and choose as they do, it is necessary to distinguish those facets of human behavior that are learned and variable from those that are innate and universal. But it is a mistake—albeit a common one—to conceptualize the problem as a search for "musical" universals. *There are none*. There are only the acoustical universals of the physical world and the bio-psychological universals of the human world. Acoustical stimuli affect the perception, cognition, and hence practice of music only through the constraining action of bio-psychological ones.[49]

These acoustical universals include basic features of our pitch perception such as octave equivalence. These are the sorts of universals that we should not be looking for in the analysis of (surface) pitch structures. These regularities and constraints are not the underlying mechanisms. To understand how these operate we need to study how the mind segments information into salient categories.

> Because of the innate capabilities of the human mind, some parameters of sound can be segmented into perceptually discrete, proportionally related stimuli that can then serve as the basis for auditory patternings. In most musics of the world, this is the case with pitch (frequency) and

---

[48] See: Leonard Meyer, *Emotion and Meaning in Music* (Chicago: University of Chicago Press, 1956); Eugene Narmour, *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*, Chapter 4 (Chicago: University of Chicago Press 1990).
[49] Meyer, "A Universe of Universals," 6.

> duration which are the basis for melody, rhythm, meter and (in Western music) harmony. Because the largely learned probabilities and possibilities that govern successions in these parameters can be the basis for syntax, I have called these parameters "syntactic".[50]

Meyer's suggestion the "syntactic" parameters are "largely learned probabilities and possibilities" accords well with Tomasello's 'pattern-finding' skill set. These domain-general capacities do much of the work too often attributed to 'universal grammar.' Still, the related idea that cognition is based on a 'particulate' system has been very influential in cognitive science and music. But it need not be allied to computational approaches alone. Combinatorial processes may operate on non-discrete elements, or 'chunks' of melodic material that are perceptually salient to particular communicative contexts. Furthermore, they may operate simultaneously to the processing of holistic elements, such as intonation curves or melodic contours. Pitch processing is therefore multidimensional and consists of several interconnected and patterned systems. There is a great deal of flexibility to our cognitive apparatus and it need not be subsumed under a single syntactocentric model. The usage-based model explains the gradience and variation that we find in melodies across cultures. It also shows how the structure of melody emerges through use as a product underlying domain-general mechanisms.

Gary Tomlinson has described several important domain-general capacities to which musical capacities are tied (and which may be compared with Tomasello's mechanisms outlined above). These include: "the ability to separate sonic events in auditory 'scene'

---

[50] Meyer, "A Universe of Universals," 8.

analysis, the grouping of proximate, related percepts into larger gestalts, abilities to time and sequence events, the interplay of procedural and episodic memory, and the astonishingly fine discrimination of timbres we deploy in all manner of auditory experience."[51] Out of these, Tomlinson finds three domains in which music may be described as special. These are entrainment, innate gesture-calls—which he links loosely to *emotive prosody*—and pitch processing. In this article Tomlinson also claims that 'discrete combinatoriality' is a universal feature of human cognition. And he considers the combinatorial component of pitch processing—which appears to be more complexly ordered than it is for linguistic use—to be *special* to music. He also speculates that it "might be an offshoot of the decoupling of the emotive, prosodic intonational patterns themselves."[52] These arguments and observations are consistent with the evidence that I have cited by Trehub and Peretz, as well as the hypotheses advanced by Charles Darwin. All take prosodic pitch patterning to be a crucial element in the human cognitive endowment for music. In Chapters 3 and 4 I will review evidence for the combinatorial principles that Tomlinson describes. I will argue that domain-general cognitive capacities account for a large measure of human musicking and that there are potentially more specialized capacities for the processing of musical pitch, whether or not it is discretized in the way generativists believe.

[51] Gary Tomlinson, "Evolutionary Studies in the Humanities: The Case of Music," *Critical Inquiry* 39 (Summer 2013): 662.
[52] Tomlinson, "Evolutionary Studies in the Humanities: The Case of Music," 666-667.

Conclusion

A biocultural musicology investigates the cognitive capacities that structure the emergence of musical thought, as well as the social mechanisms by which specific musical practices are learned. This is best understood as a usage-based approach, and it is supported by the kinds of empirical studies surveyed in this chapter, including important work on music and language acquisition, the cognitive deficits associated with amusia, brain localization studies, comparative and cross-cultural analysis of music in world cultures, and studies of music and evolution. The usage-based approach opens up new perspectives on song melody. For instance, the features of pitch patterning in song may be understood as the product of social learning and usage rather than the innate, biologically determined propensities proposed by rationalists. I have shown that there is no significant evidence to support the nativist hypothesis. Music is best understood as a complex adaptive system that is the product of multiple interacting forces through phylogeny. The prosodic features and combinatorial elements shared by music and language may be explained by the action of domain-general cognitive capacities in conjunction with social-cognitive learning mechanisms. Grammaticalization takes place through usage. Social learning plays a crucial role in our acquisition of, and subsequent proficiency for, music and language. The usage-based framework sketched here—and developed and applied in more detail through the dissertation—is interdisciplinary in scope, and seeks to counter the ethnocentrism of much contemporary music cognition by providing a rigorous comparative framework for studying music in world cultures.

CHAPTER 2

Rationalism and empiricism in the study of language, music, and mind

There is an extraordinary diversity of pitch patterning in song both within and across cultures. How do listeners make sense of this variation? What are the shared cognitive resources that enable us to adapt to tonal systems that employ very different categories and modes of organization? Are these resources innate or do we learn from experience?[53] Debates framing questions such as these have turned on rationalist as opposed to empiricist epistemologies.[54] Few *current* music cognition studies have engaged in these

---

[53] "At one extreme, the capacity for acquiring musical abilities is seen as an evolutionary adaptation, shaped by natural selection and governed by genes. At the other extreme, musical abilities are viewed as the result of general-purpose learning capacities that are shaped by the environment – the 'blank slate' or the tabula rasa scenario" (Aniruddh Patel, *Music, Language and the Brain* (New York: Oxford, 2008): 3.

[54] I recognize that there is a rationalist element to the arguments put forward in this dissertation. What I signal here is not antipathy to all aspects of rationalist inquiry per se. This chapter addresses specific positions in the literature that are inconsistent with what we know of music in world cultures, with the evidence surveyed in Chapter 1, and with the empirical data presented in Chapter 4. It is also important to recognize that debates involving rationalism versus empiricism have a long history in philosophy, even if these are of a different tenor to the current impasse: "In discussions of the history of philosophy, it is common to talk of a showdown in the seventeenth and eighteenth centuries between 'the rationalists' and the 'empiricists'. Rationalists like Descartes and Leibniz believed that pure reasoning can be a route to knowledge that does not depend on experience. Mathematics seemed to be a compelling example of this kind of knowledge. Empiricists like Locke and Hume insisted that experience is our only way of finding out what the world is like. In the late eighteenth century, a sophisticated intermediate position was developed by the German philosopher Immanuel Kant. Kant argued that all our thinking involves a subtle *interaction* between experience and preexisting mental structures that we use to *make sense* of experience. Key concepts like space, time, and causation cannot be derived from experience, because a person must *already* have these concepts in order to use experience to learn about the world. Kant also held that mathematics gives us real knowledge of the world but does not require experience for its justification" (Peter Godfrey-Smith, *Theory and Reality: An Introduction to the Philosophy of Science* (Chicago: University of Chicago Press, 2003): 20-21). Clearly, these are not the same debates that figure in contemporary cognitive science. I have use the terms rationalism and empiricism not as practiced by Descartes, Locke and other philosophers, but as very broad classifications for contemporary approaches to

debates even though they are central to all work in cognitive science. This chapter takes this to be a priority. In order for us to make headway on basic problems in the study of music and mind we must reflect on the assumptions that inform our research. At present, this means taking a hard look at the rationalist modes of inquiry that have dominated the field for several decades now. Some of these were discussed in Chapter 1 and include elements of what Jerry Fodor has described as the 'New Synthesis', a fusion of computationism, modularity, nativism, and evolutionary psychology.[55] The rationalist epistemology of current generative linguistics and music theory remains the guiding framework despite its deficiencies. In this first part of the chapter I unpack several key aspects of rationalist thought that restrict investigations of music and mind prematurely. Later on in the chapter I develop the alternative position: a multi-faceted and largely empiricist perspective that treat song as a *complex adaptive system* and that builds on the usage-based approach discussed at the end of Chapter 1.

The first part of this chapter offers a detailed critique of rationalist inquiry in music and linguistic theory. To understand the role of generative theory we must look to its origins in linguistics. Many of the assumptions outlined in Chomsky's early work remain entrenched in current thinking, including various instantiations of the generative theory of tonal music (GTTM) developed initially by Fred Lerdahl and Ray Jackendoff (1983).

---

the study of music, language and mind. I use these terms to point out the strengths and weaknesses of these contrasting approaches, but in doing so I recognize that the boundaries are not hard and fast. For instance, 'empiricists' also propose certain 'innate' mechanisms (it is just that these are not taken to be special-purpose).

[55] Jerry Fodor, *The Mind Doesn't Work That Way* (Cambridge, MA: MIT Press, 2001).

PART 1: GENERATIVE APPROACHES

2.1 Generative Theory of Tonal Music

Fred Lerdahl and Ray Jackendoff's *Generative Theory of Tonal Music* (1983) was the first system analogous to that of generative-transformational grammar to be developed in music theory. [56] In this foundational text they relied on many of the concepts outlined by Noam Chomsky in several classic monographs (Chomsky 1965, 1968, 1975).[57] But theirs was not a simplistic 'application' of linguistic theory to music theory, as Lerdahl points out his more recent overview of GTTM:

> Our interest was not in a literal transfer of linguistic to musical concepts, as Leonard Bernstein (1976) attempted. Rather, it was Chomsky's way of framing issues that attracted us: the supposition of specialized mental capacities, the belief that they could be studied rigorously by investigating the structure of their outputs, the distinction between an idealized capacity and its external and often accidental manifestations, the idea of a limited set of principles or rules that could generate a potentially infinite set of outputs, and the possibility that some of these principles might be unvarying beneath a capacity's many different cultural manifestations.[58]

Our concerns in the first half of this chapter are less in the details of GTTM's preference rules and prolongational structures and more in the epistemology embodied in its generative framework; that is, in several of the ideas outlined in this very extract. We need to investigate the assumptions that guide the theory because these have gained wide

---

[56] Fred Lerdahl and Ray Jackendoff, *A Generative Theory of Tonal Music* (New York: Oxford, 1983).
[57] Noam Chomsky, *Aspects of the Theory of Syntax* (Cambridge, MA: MIT Press, 1965); Noam Chomsky, *Language and Mind* (New York: Oxford, 1967); Noam Chomsky, *Reflections on Language* (New York, NY: Pantheon, 1975).
[58] Fred Lerdahl, "Genesis and Architecture of the GTTM Project," *Music Perception* 26, No. 3 (2009): 187.

acceptance in music cognition over the past three decades, and well beyond the reach of GTTM itself.[59]

GTTM proposes a rationalist, nativist, mentalistic approach to cognition that hypothesizes a 'music acquisition device' very much akin to the 'language acquisition device' introduced in Chapter 1: "The musical capacity constitutes the resources in the human mind/brain that make it possible for a human to acquire the ability to understand music in any of the musical idioms of the world, given the appropriate input."[60] A distinction is made between the "*broad musical capacity*, which includes any aspect of the mind/brain involved in the acquisition and processing of music, and the *narrow musical capacity*, which includes just those aspects that are specific to music and play no role in other cognitive activities."[61] Crucially, music is understood to have domain-specific characteristics that distinguish it from language. The focus is on what makes music unique rather than on how it fits into the general human cognitive makeup, and GTTM tackles these issues from a special vantage point. Lerdahl and Jackendoff "take the goal of a theory of music to be a *formal description of the musical intuitions of a listener who is experienced in a musical idiom*."[62] A formal description of this sort must

---

[59] Since we are concerned primarily with pitch patterning in this study it is worth pointing out that the prolongational features of GTTM are not well suited to the analysis of melody. To some extent this was the rationale for Lerdahl's more recent *Tonal Pitch Space* (2001) in which he formulated more sophisticated rules that circumscribe melodic and harmonic features of Western tonal music, and, to some extent, of later chromatic developments. The over-arching theoretical framework is nevertheless intact and will be discussed below (Fred Lerdahl, *Tonal Pitch Space* (New York: Oxford, 2001).
[60] Ray Jackendoff and Fred Lerdahl, "The Capacity for Music: What Is It, And What's Special About It?" *Cognition* 100 (2006): 35.
[61] Ibid.
[62] Lerdahl and Jackendoff, *A Generative Theory of Tonal Music*, 1.

propose context-free rules that operate across cultures. As we shall see, GTTM ignores the "external and often accidental manifestations" that we find in many performances. Instead it must invoke certain idealizations in order to account for the intuitions it claims are universal. These 'musical universals' "need not be readily apparent in all musical idioms. We are concerned here […] with universals of musical *grammars*—the principles available to all experienced listeners for organizing the musical surfaces they hear, no matter what idiom they are experienced in."[63] Several other idealizations are outlined by Lerdahl in his overview in *Tonal Pitch Space* (2001):

> GTTM makes assumptions analogous to those found elsewhere in cognitive science. First, it takes as given the musical surface (the aural perception of pitches, timbres, durations, and dynamics), ignoring the complex process by which the surface is constructed from the acoustic signal. In the same spirit, the notated score is taken—minus bar lines but with the addition of harmonic roots and tonal orientations—to represent the surface. Second, GTTM assumes an "experienced listener." In reality no two listeners are exactly alike, nor are any two hearings by the same listener. Given familiarity with an idiom, however, the ways in which a piece is understood are highly constrained. A theory of musical understanding needs to characterize these common constraints as a framework for studying individual differences. Third, the theory provides structural descriptions not for how the music is heard as it unfolds in time but for the final state of a listener's understanding. *It is doubtful whether a substantive theory of real-time listening processes can be developed without first considering the nature of the information these processes deliver*. […] A fourth limitation concerns the distinction between hierarchical and associational structures (Meyer 1973). By "hierarchy" is meant an organization of discrete elements such that

[63] Lerdahl and Jackendoff, *A Generative Theory of Tonal Music*, 278.

> one element is perceived to subsume or contain other elements. The elements do not overlap, subsumption is recursive, and at any given level the elements are adjacent. By "association" is meant the relative closeness or distance that is judged to exist between ideas. An idea is closely associated with another if it takes few psychologically viable operations to change one into the other. […] A fifth idealization in the GTTM theory is the construal of a musical surface as a single sequence of discrete events that assemble into hierarchically organized groupings [emphasis added].[64]

Are these assumptions and idealizations justified in light of cross-cultural evidence for human musicking? Should we accept models of Western music and listening as interchangeable with all others? Many of the ideas expressed in this extract remain contested. Is there 'cognitive reality' to discrete, hierarchically ordered elements? Should we eliminate real-time performances from our analysis, and if so, why? Should we limit our analysis to the 'musical surface'? Should we privilege the 'experienced listener' over the 'listener,' or even the performer instead? A reflexive cognitive music theory must seek answers to all these questions without making the sorts of assumptions that Lerdahl and Jackendoff take to be unproblematic. The next four sections of this chapter form an extended critique of these rationalist assumptions and what they mean for a theory of music cognition.

---

[64] Fred Lerdahl, *Tonal Pitch Space*, 6-7.

2.2 Notation and the musical surface

The musical score is the basis for all analyses in generative music theory because music notation provides a convenient objectification of the 'musical surface.' This idea of taking the 'musical surface' as *given* assumes that the appropriate object of inquiry is 'musical competence,' or the knowledge of music that exists in the mind of the listener. Generativists argue that *direct* evidence of musical behavior (from actual performances) is misleading for music cognition research because it includes a great deal that is simply accidental and superfluous. I will argue precisely the opposite: that in studying the production and perception of song sounds we must attend to details of performance that direct attention to relevant communicative features. These features are often eliminated from notated transcriptions because they do not conform to conventional categories of staff notation, or they are included as ornamental features (like glissandi, for instance) that are not categorical in their own right.

The score-based paradigm is not unique to generative music theory. It is indicative of a wider practice in formalist music theory, including the Schenkerian framework upon which much of GTTM is based. Lerdahl and Jackendoff claim that, "the domain of formal analysis lends itself best to exploring the full richness and complexity of musical understanding."[65] But by 'musical understanding' they mean of course 'musical competence.' The 'object' of this analysis should be the "*musical surface*: the array of

---

[65] Jackendoff and Lerdahl, "The Capacity for Music: What Is It, And What's Special About It?" 35.

simultaneous and sequential sounds with pitch, timbre, intensity, and duration."[66] The musical score is the visual representation of this surface and it is Western tonal music (classic and popular) that is used almost exclusively for analysis of this sort.[67] Because the notational features of Western tonal music are both cognitive and descriptive they provide a robust system for the analysis of several parameters, especially pitch and duration. Yet despite these advantages, there are also significant drawbacks to using the musical score. We need to recognize that the discrete categories of staff notation are culture-specific tools that are not equivalent to properties of mind.[68] The categories and functions delineated by staff notation map to very few other cultures. In many cases they exclude important categories of pitch structure, such as the contours and glides that I describe in Chapter 4 on Zulu song prosody. In sum, a cognitive music theory that is dependent on the categories and relations of staff notation cannot account for many important features of Western and non-Western musics alike. Generative theorists tend to

---

[66] Ibid, 37. "Beyond the musical surface, structure is built out of the confluence of two independent hierarchical dimensions of organization: rhythm and pitch. In turn, rhythmic organization is the product of two independent hierarchical structures, grouping and meter. The relative independence of rhythmic and pitch structures is indicated by the possibility of dissociating them. […] each of these components has its own characteristic units and combinatorial principles" (Ibid, 37).

[67] It is important to recognize that important details of musical performance are lost in staff notation whether or not the music notated is Western. However, the 'surplus' of non-discrete information is greater for many non-Western music cultures that do not conform to the categories it codifies. See: Lerdahl and Jackendoff, *A Generative Theory of Tonal Music*; Lerdahl, *Tonal Pitch Space*; Jackendoff and Lerdahl, "The Capacity for Music: What Is It, And What's Special About It?"

[68] "[We] lose sight of the need for explanation when phenomena are too familiar and 'obvious.' We tend too easily to assume that explanations must be transparent and close to the surface. The greatest defect of classical philosophy of mind, both rationalist and empiricist, seems to me to be its unquestioned assumption that the properties and content of the mind are accessible to introspection." Noam Chomsky, *Language and Mind*, Third Edition (Cambridge: Cambridge University Press, 2006): 22.

overlook these deficiencies by focusing on a very limited repertoire and by hypothesizing an innate capacity.[69]

This ethnocentrism is not unique to GTTM. The grammar of Western common practice tonal music has been used as the basis for many experimental studies. Carol Krumhansl's influential *Cognitive Foundations of Musical Pitch* (2001), for instance, extrapolates on the insights of cognitive music theorists who also use this repertory almost exclusively.[70] Discrete and absolute pitch categories and intervallic structures are the basis for many of her experiments, including her diagnostic probe tone tests. David Huron has based several experiments on the notated features of Western tonal music as well as the contrapuntal elements of its grammar. It is difficult to see how the results of these studies, insightful though they are, can usefully be generalized as the basis for tonality across cultures.[71] Huron himself points to such limitations with the use of notated structures:

> In music theory, naïve realism is evident in two assumptions: that the structures we see in the notated music are the ones we experience, and that the structures we experience can be seen in the notation. Those who aim to describe the music as it is can do so only by abandoning the path of naïve realism. Sophisticated realism begins by acknowledging the separation between the

---

[69] "The grammar that music theory teaches is unavoidably tied to the repertoire to which it refers, and just how this is generalized to apply to other repertoires is not immediately apparent." Lawrence Zbikowski, *Conceptualizing Music: Cognitive structure, theory and analysis*, AMS Studies in Music (New York: Oxford, 2003): ix.

[70] Carol Krumhansl, *Cognitive Foundations of Musical Pitch*, Second Edition (New York: Oxford, 2001).

[71] David Huron, "Interval-class content in equally-tempered pitch-class sets: Common scales exhibit optimum tonal consonance," *Music Perception* 11, no. 3 (1994): 289-305; David Huron, "Tone and voice: A derivation of voice-leading from perceptual principles," *Music Perception* 19, no. 1 (2001): 1-64.

> subjective and objective worlds. Sophisticated realism is possible only by making use of empirical methods of observation.[72]

We should not assume that cognitive functions are embodied in notational conventions, nor should we assume that the "properties and content of mind are open to introspection." We need to recognize that notation determines not only the performance and composition of music, but also its analysis. The idea that notation *is* music remains a powerful ideology in cognitive music theory. It is a problem that exists on several levels. First, the acoustic data of the auditory signal (including both song and speech) are not easily reduced to the categories of staff notation and cannot be interpreted in a simplistic way as indicative of competence. A 'musical surface' of the sort envisaged by Jackendoff and Lerdahl misses out on gradient elements of pitch patterning that cannot be rendered in staff notation. Notation is a script for *idealized* performances, not a transcript of the processes engendered by *actual* performances. It provides only indirect evidence of *select* processes. Notation is a tool that mediates our experience of sound and should not be mistaken for the facts of production and perception. Again, it is in no sense an accurate symbolic rendering of the acoustic signal, nor the processes that engender it.

2.3 The experienced listener

We are all expert listeners. All normal humans are capable of understanding and responding to music regardless of 'training' because we acquire these skills through ontogeny. What Lerdahl and Jackendoff have in mind, then, when they refer to the

[72] Huron, *Sweet Anticipation*, 371-2.

'experienced listener,' is something different. It is an idea that dates back to early cognitive music theory in the United States and it is tethered to a particular role accorded the analyst. "As I intend the term, criticism seeks to explain how the structure and process of a particular composition are related to the competent listener's comprehension of it. In other words, the role of the music critic is similar to that of the literary critic."[73] Leonard Meyer's usage of the term, like that of Lerdahl and Jackendoff, is predicated on a system of value that was a product of the mid-twentieth century American academy. Meyer was explicit about the aims of his analysis and the empirical ground from which it emerged. In *Explaining Music* (1973), he describes what he takes music analysis to be:

> Values are, of course, always latent in what the critic does—in his account of the ways musical relationships affect the listener's understanding of and response to particular musical works, and also in his choice of works to analyze. But the critic does not, I think, begin with aesthetic principles and arrive at critical judgments. Quite the opposite. He begins with his own responses—his cognitive-affective sense of whether a composition is convincing and exciting, intriguing and entertaining. Then he attempts to find rational grounds for his judgment. […] Moreover, since he deals for the most part with works by acknowledged masters, it seems a bit pretentious for the critic to take as his main task and examination of the virtues of compositions by Bach or Beethoven, Haydn or Handel. To do so is probably also circular, and perhaps somewhat disingenuous. For the works of such masters are in some sense the initial basis for his stylistic standard and his criteria of value. In short, the critic does not come to praise masterpieces, but to explicate and illuminate them: to understand and explain, that is, how the various sorts of tonal

[73] Leonard Meyer, *Explaining Music: Essays and Explorations* (Chicago: Chicago University Press, 1973): ix.

> relationships in a particular composition are understood and enjoyed by experienced sensitive listeners.[74]

In many ways GTTM works within this tradition: it focuses on "works by acknowledged masters" and how these are "enjoyed by experienced sensitive listeners." Meyer was a humanist for whom intuition played an important role in analysis. In this he is part of a long line of music theorists for whom the act and art of analysis are themselves a performance. Similarly, Lerdahl writes that: "our quest for cognitive principles would proceed from our own musical intuitions. Only later would we seek experimental corroboration."[75] Intuition may serve the critic well, but there is no reason why it should guide our investigations from the start, or at least not in studies of music perception and cognition where a measure of distanciation and objectivity are necessary for unbiased results. Another point about this focus on experienced listeners is that it assumes a particular repertory; in other words, it relays the work-centric focus of music theory to music cognition. This manifests itself in the third idealization made by Lerdahl and Jackendoff: the hypothesis that structural descriptions of the final state are not real-time.

2.4 The 'final state' vs. processual models

> [W]e would build a final-state rather than processing theory, on the view that it was advantageous to specify the mental structures in question before trying to articulate how they operated in real

[74] Meyer, *Explaining Music*, ix-x.
[75] Lerdahl, "Genesis and Architecture of the GTTM Project," 188.

> time. These positions were not meant to denigrate psychoacoustics, experiment, or processing. Rather, they were strategic decisions in theory construction. They also had the advantage of keeping our project within the bounds of the music-theory tradition, since most music theories take pitches and rhythms for granted, appeal to intuitive plausibility, and do not restrict the study of musical works to their real-time unfolding.[76]

Keeping the project "within the bounds of the music-theory tradition" may not be the most effective strategy precisely because of its reliance on the subjective responses of the analyst. We also need also be skeptical of final-state knowledge. By fixing the musical object in time and space and then reducing its features to a series of processes and hierarchies, real-time cognitive processes are ignored and the details of actual performance data are eliminated. This emphasis on form and structure at the expense of process takes for granted a work-based music theory. It also reinforces the idea, borrowed from generative linguistics, that 'competence' rather than 'performance' is the proper domain of analysis.

An important tenet of the usage-based music theory developed in this dissertation is that music is constructed through acts of performance. We need to understand music as process, not simply as product. Edward T. Cone made a useful distinction between 'synoptic comprehension' and 'immediate apprehension' to characterize two modes of music perception.[77] The former is directed at questions of unity and structure while the

[76] Lerdahl, "Genesis and Architecture of the GTTM Project," 188.
[77] For a discussion of these issues see: Kofi Agawu, "How We Got Out of Analysis and How to Get Back in Again," *Music Analysis* 23, Nos. 2-3 (2004): 267-286.

latter is “the mode by which we directly perceive the sensuous medium, its primitive elements, and their closest interrelationships.”[78]

> Immediate apprehension is our response to a direct contact – our recognition of Whitehead’s ‘vivid values.’ Synoptic comprehension is to some extent conceptual: it is our realization of the form of what we have perceived – Whitehead’s ‘organism.’ […] Of the two modes, it is the immediate that enjoys both temporal and logical priority in our perception of art. Temporal, in the sense that our appreciation of an esthetic object usually begins with our apprehension of its sensuous qualities and, especially in the case of a time-art, of its details; logical, because, in my view, enjoyment of such apprehension can lead to some measure of esthetic satisfaction whether or not it is accompanied by synoptic comprehension, and whether or not such comprehension, if achieved, finds a worthy object.[79]

If the ‘immediate,’ as Cone describes it, “enjoys both temporal and logical priority in the perception of art” then why should it play a secondary role in theories of music? While it would be a mistake to argue that composers are not concerned with “the constructive or organizing element in music,” it is equally problematic to assume this to be their principle goal, or, more importantly, that it is composers and works that should be the sole focus of our investigations. The degree to which perception and cognition is synoptic is limited not only by restrictions on memory, but also by the material itself. Many songs are short and repetitive and do not consist of the extended sentences, paragraphs and movements of the symphonies and sonatas taken as normative. In other words, *tonal*

---

[78] James Webster, “*Formenlehre* in Theory and Practice,” in *Musical Form, Forms, and Formenlehre: Three Methodological Reflections*, ed. by Pieter Bergé (Leuven: Leuven University Press, 2009): 124.
[79] Edward T. Cone, *Musical form and musical performance* (New York: Norton, 1968): 88-90.

development is not a universal feature of musical structure (see Chapter 4, Part 3) and so our theories of music perception and cognition should not assume this to be the case. Whereas GTTM and many other music theories are composer- and work-centric, the usage-based approach adopted in this dissertation seeks a more balanced appraisal of both production and perception.

Texts communicate information and mediate acts of interpretation and perception. This explains why the implementation of formal models cannot do justice to musical experience. What Cone's observations suggest is that narrative description, intuitive deduction, and analysis based on generic formal templates hinder our efforts at achieving a scientific basis for cognitive music theory.[80] In the study of music and mind *we cannot take the products of compositional invention as natural*. The conventions of Western art music were formulated and codified over millennia. These are learned, culture-specific phenomena. While they may display properties and propensities indicative of human musicking as a whole, we should be cautious in attributing this primarily to the syntactic features of a codified style. Syntactocentrism in both linguistics and music tends to ignore prosodic and other features of which melody is a part.

2.5 Hierarchical structure

The final two idealizations outlined by Lerdahl apply to the organization and structure of the musical surface. The assumption here is that the basic units of these hierarchies are

[80] By 'formal templates' I mean generic music-theoretical items like 'sonata form' and 'rondo' rather than 'mental' templates in general.

discrete. We see this in the fourth idealization: "By 'hierarchy' is meant an organization of discrete elements such that one element is perceived to subsume or contain other elements. […] A fifth idealization in the GTTM theory is the construal of a musical surface as a single sequence of discrete events that assemble into hierarchically organized groupings."[81] But is it really feasible to segment the musical surface into discrete events of this sort? Are such events as self-evident as the musical score makes out? In Chapter 4 I present evidence to show that pitch categories consist of non-discrete elements and possess gradient features. This gradience is not arbitrary and is in fact crucial to their operation. These categories may include minimal elements of pitch information as well as larger 'chunks.' In other words, the 'discrete' (pitch and durational) events depicted in the musical score (musical surface) are not consistent with the evidence we now have for the categories of perception and cognition (whose consistency is entirely contextual and has nothing to do with their discreteness). To understand why GTTM ignores these and many other features discussed in Part 1 we must turn to a more detailed discussion of Chomsky's Cartesian linguistics and the rationalist epistemology that guides it.

PART 2: RATIONALISM

2.6 Cartesian Linguistics: Chomsky and Rationalism

The most influential figure in cognitive science and linguistics to have championed the rationalist cause is undoubtedly Noam Chomsky. In *Syntactic Structures* (1957), *Aspects of the theory of syntax* (1965), and especially *Cartesian Linguistics* (1966) he outlined a

[81] Lerdahl, *Tonal Pitch Space*, 6-7.

theory of language based on rationalist and romantic philosophers from Descartes to Humboldt as well as new work in mathematics, symbolic logic, and computation. He and his colleagues at MIT developed a rigorous and coherent research program that, in its various forms, has dominated the field of linguistics over the past four decades and which is relevant to many of the linguistic theories discussed in this dissertation. Generative linguistics has also exerted a profound influence on theories of cognitive science and music. Several concepts have gained currency, including those of transformational generative grammar, discrete combinatoriality, knowledge of language, and the more recent minimalist program. The core issue that binds all of these and which is central to the 'biological' study of language and music is *nativism*. Chomsky's allegiance to nativism, and the universal attributes associated with it, is unusual in its specificity. In *Cartesian Linguistics* he recognizes historical precedents for his approach.[82]

> The central doctrine of Cartesian linguistics is that the general features of grammatical structure are common to all languages and reflect certain fundamental properties of the mind. It is this assumption which led the philosophical grammarians to concentrate on *grammaire généraleé*

[82] Chomsky makes it clear that his reading of Cartesian linguistics is idiosyncratic and limited: "I will make no attempt to characterize Cartesian linguistics as it saw itself, but rather will concentrate on the development of ideas that have re-emerged, quite independently, in current work. My primary aim is simply to bring to the attention of those involved in the study of generative grammar and its implications some of the little-known work which has bearing on their concerns and problems and which often anticipates some of their specific conclusions" (Chomsky 1966, 58). Still, there are several features of Cartesian thought that are foundational to the generative program, especially the 'creative principle' that Chomsky takes as central to human action: "In summary, it is the diversity of human behavior, its appropriateness to new situations, and man's capacity to innovate – the creative aspect of language use providing the principal indication of this – that leads Descartes to attribute possession of mind to other humans, since he regards this capacity as beyond the limitations of any imaginable mechanism. Thus a fully adequate psychology requires the postulation of a 'creative principle' alongside of the 'mechanical principle' that suffices to account for all other aspects of the inanimate and animate world and for a significant range of human actions and 'passions' as well." Noam Chomsky, *Cartesian Linguistics: A Chapter in the History of Rationalist Thought* (New York: Harper and Row, 1966): 61.

> rather than *grammaire particulière* and which expresses itself in Humboldt's belief that deep analysis will show a common "form of language" underlying national and individual variety. There are, then, certain language universals that set limits to the variety of human language. The study of the universal conditions that prescribe the form of any human language is "grammaire générale." Such universal conditions are not learned; rather, they provide the organizing principles that make language learning possible, that must exist if data is to lead to knowledge. By attributing such principles to the mind, as an innate property, it becomes possible to account for the quite obvious fact that the speaker of a language knows a great deal that he has not learned.[83]

Chomsky claims that the universal conditions and organizing principles that he attributes to the mind are innate. He would later term this 'knowledge of language,' or the general organizing principles, 'universal grammar' or UG.[84] Particular grammars, on the other hand, are those rules specific to one or another language. The interplay of these two grammars enabled Chomsky to postulate his rationalist theory of language without having to explain the nature of linguistic diversity and change. But research on language acquisition since the 1980s has accumulated considerable evidence to show that social learning accounts for a great deal more than Chomsky initially supposed (Tomasello 2003: Chapter 1). This new evidence undermines the distinction Chomsky makes between 'universal' and 'particular' grammars by showing its emergence to be the product of specific social contexts and inheritances. The idea that underlying all

---

[83] Chomsky, *Cartesian Linguistics*, 59-60.

[84] Noam Chomsky, *Syntactic Structures* (The Hague: Mouton, 1957). "Chomsky had the revolutionary idea of construing the grammar of a natural language as equivalent to an automaton capable of generating the sentences of that language – a *generative grammar* – and asked what sort of grammar would be adequate." Adele Abrahamsen and William Bechtel, "History and Core Themes," in *The Cambridge Handbook of Cognitive Science*, ed. by Keith Frankish and William M. Ramsey (Cambridge, Cambridge University Press, 2012): 14.

languages are similar "organizing principles that make language learning possible" is now more difficult to sustain.

Chomsky signaled his opposition to empiricism most vociferously in his famous review of B.F. Skinner's *Verbal Behavior* in which he argued that principles of association and induction are insufficient to guide language acquisition.[85] Instead he proposed that there are innate mechanisms that guide this process.[86] This nativist program with its rationalist epistemology came to exert a powerful influence on the formation of the cognitive sciences.[87] Chomsky's main argument in support of nativism is the *poverty of the stimulus*, the idea that the learning environment cannot account for the complexity of language acquisition.

> On the basis of the best information now available, it seems reasonable to suppose that a child cannot help constructing a particular sort of transformational grammar to account for the data

[85] Noam Chomsky, "Verbal Behavior by B.F. Skinner," *Language* 35, No. 1 (1959): 26-58.
[86] "Perhaps more than any other, the issue of innateness divides rationalist and empiricist approaches to the science of cognition. Rationalist and empiricist alike believe that there are innate cognitive capacities. Everyone believes we are born with the capacity to learn. What divides rationalist from empiricist is the idea that there is innate knowledge; that substantive and contingent information about the environment is part of the genetic endowment of every biological cognizer. The central argument for this claim is very simple and goes back to Plato: certain things that are known could not be learned and hence must be innate." Robert Cummins and Denise Dellarosa Cummins, eds., *Minds, Brains, and Computers: The Foundations of Cognitive Science* (Malden, MA: Blackwell, 2000): 447.
[87] "Between about 1960 and 1985, most of linguistics, psychology, artificial intelligence, and natural language processing was completely dominated by a *rationalist* approach. A rationalist approach is characterized by the belief that a significant part of the knowledge in the human mind is not derived by the senses but is fixed in advance, presumably by genetic inheritance. Within linguistics, this rationalist position has come to dominate the field due to the widespread acceptance of arguments by Noam Chomsky for an innate language faculty. Within artificial intelligence, rationalist beliefs can be seen as supporting the attempt to create intelligent systems by handcoding into them a lot of starting knowledge and reasoning mechanisms, so as to duplicate what the human brain begins with." Christopher D. Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing* (Cambridge, MA: MIT Press, 1999): 4-5.

> presented to him, any more than he can control his perception of solid objects or his attention to line and angle. Thus it may well be that the general features of language structure reflect, not so much the course of one's experience, but rather the general character of one's capacity to acquire knowledge—in the traditional sense, one's innate ideas and innate principles.[88]

The generative enterprise teeters on these claims, and, if empiricists are able to establish that even a fraction of linguistic operations depend on domain-general mechanisms rather than "innate ideas" and "principles," then it may undermine the nativist program entirely. This program is explicitly opposed to that of empiricism, and so we need compare and contrast these two approaches.

Nativism is a crucial and definitive difference, but it is not absolute since empiricists do agree that there are some domain-general cognitive capacities present at birth. What they contest is the idea that there is specialist knowledge: for empiricists there are no pre-determined rules, operations, or other mental procedures *specific* to language or music processing that are the product of biologically encoded modules in the brain. Rather, the acquisition of language and music is dependent on general-purpose learning mechanisms that are applied to the child's learning environment through ontogeny. Acquisition is an incremental social process that takes place through the interaction of mind/brain and environment.

---

[88] Chomsky, *Aspects of the Theory of Syntax*, 59. "It is, for the present, impossible to formulate an assumption about initial, innate structure rich enough to account for the fact that grammatical knowledge is attained on the basis of the evidence available to the learner. Consequently, the empiricist effort to show how the assumptions about a language-acquisition device can be *reduced to a conceptual minimum* is quite misplaced. The real problem is that of developing a hypothesis about initial structure that is sufficiently rich to account for acquisition of language, yet not so rich as to be inconsistent with the known diversity of language" (Ibid, 57).

A second crucial difference between rationalists and empiricists has to do with the sorts of data they collect and study. Empiricists are interested in *direct* evidence of *language usage* whereas rationalists require only *indirect* evidence of competence (in music studies the latter is equated with the 'musical surface'). Direct evidence usually means recordings of actual speech that are studied using spectrogram software or other methods of annotation and analysis, including corpus studies. *Indirect* evidence tends to involve the analysis of texts as linguistic objects and manifests itself in formalist studies that propose abstract rules for cognitive operations or procedures.[89] This focus on texts and knowledge of language is linked to a fundamental distinction that Chomsky makes between 'competence' and 'performance,' or what he later characterized as 'internal' (or 'I-language') versus 'external' (or 'E-language') forms. As Cummins and Cummins write in describing Chomsky's approach:

> Rationalists and empiricists are attempting to describe the language module of the human mind (the I-language) for which data such as texts (the E-language) provide only indirect evidence, which can be supplemented by native speaker intuitions. Empiricist approaches are interested in describing the E-language as it actually occurs. Chomsky (1965: 3-4) thus makes a crucial distinction between *linguistic competence*, which reflects the knowledge of language structure that is assumed to be in the mind of a native speaker, and *linguistic performance* in the world, which is affected by all sorts of things such as memory limitations and distracting noises in the environment. Generative linguistics has argued that one can isolate linguistic competence and

[89] Modern music theory has given priority to the study of scores, notation, and musical works. When the cognitive sciences of music were established this bias was inherited whether or not a rationalist or an empiricist epistemology was assumed.

> describe it in isolation, while empiricist approaches generally reject this notion and want to describe actual use of language.[90]

The idea of isolating linguistic competence rather than studying usage is a basic methodological difference. But why has the study of 'performance' been so neglected by cognitivists?

In his influential *Cours de linguistique générale*, Ferdinand de Saussure argued that speech be omitted from the study of linguistics and that we study an idealization of linguistic structure instead.[91] Saussure, like Chomsky after him, regarded the study of phonetics and speech variation to be outside the domain of formal linguistics. He argued that 'language,' as a social phenomenon, exists independently of phonetic changes and that it is a 'fixed' system. Chomsky's slightly different articulation of this distinction is outlined in the influential *Sound Pattern of English* (Chomsky and Halle 1968):

---

[90] Cummins and Cummins, *Minds, Brains, and Computers*, 6.

[91] Saussure went to great lengths to dismiss 'phonetic changes' and the 'mistakes of performance' as beyond the realm of linguistics: "Language is comparable to a symphony in that what the symphony actually is stands completely apart from how it is performed; the mistakes that musicians make in playing the symphony do not compromise this fact. An argument against separating phonation from language might be phonetic changes, the alterations of the sounds which occur in speaking and which exert such a profound influence on the future of language itself. Do we really have the right to pretend that language exists independently of phonetic changes? Yes, for they affect only the material substance of words. If they attack language as a system of signs, it is only indirectly, through subsequent changes of interpretation; there is nothing phonetic in the phenomenon. Determining the causes of phonetic changes may be of interest, and the study of sounds will be helpful on this point; not none of this is essential: in the science of language, all we need to do is to observe the transformations of sounds and to calculate their effects. What I have said about phonation applies to all other parts of speaking. The activity of the speaker should be studied in a number of disciplines which have no place in linguistics except through their relation to language." Ferdinand de Saussure, *Course in General Linguistics*, translated by Wade Baskin, ed. by Perry Meisel and Haun Saussy (New York: Columbia University Press, 2011): 18.

> The speaker produces a signal with a certain intended meaning; the hearer receives a signal and attempts to determine what was said and what was intended. The performance of the speaker or hearer is a complex matter that involves many factors. One fundamental factor involved in the speaker-hearer's performance is his knowledge of the grammar that determines an intrinsic connection of sound and meaning for each sentence. We refer to this knowledge—for the most part, obviously, unconscious knowledge—as the speaker-hearer's "competence." Competence, in this sense, is not to be confused with performance. Performance, that is, what the speaker-hearer actually does, is based not only on his knowledge of the language, but on many other factors as well—factors such as memory restrictions, inattention, distraction, nonlinguistic knowledge and beliefs, and so on. We may, if we like, think of the study of competence as the study of the potential performance of an idealized speaker-hearer who is unaffected by such grammatically irrelevant factors. […] The person who has acquired knowledge of a language has internalized a system of rules that determines sound-meaning connections for indefinitely many sentences.[92]

What Chomsky describes as 'knowledge of language' is innately acquired. It is, in a fundamental sense, *disembodied knowledge*, and this is what constitutes 'competence.'[93] As Gregory Guy writes,

> One of the central elements of the conceptual framework underlying most of modern linguistics is a fundamental opposition between, first, what is considered the essential *system* of language – the abstract mental construct of processes and elements that define what is possible in a language and comprise the generative capacity of a speaker – and second, the operations and products of that

[92] Noam Chomsky and Morris Halle, *The Sound Pattern of English* (1968, Repr. Cambridge, MA: MIT Press, 1991): 3.

[93] From the standpoint of embodied music cognition we may argue that the very potential for language is shaped by its mode of realization and not simply by an abstract knowledge system. This perspective assumes a holistic cognitive organism and not one whose behavior is determined by innate mechanisms.

> system, the actual *usage* of language by speakers. This opposition has been formulated in several slightly differing ways, so that Saussure (1922) enunciates a distinction between *langue* and *parole*, while Chomsky (1965) contrasts competence and performance. But the elements of the distinction remain the same. On the one hand there is the abstract, not directly observable construct: the grammar, *langue*, competence; and on the other hand there is the concrete, observable sum of language production: *parole*, performance, utterances.[94]

For generativists, knowledge of language is the product of a mentalistic grammar that is immune to the influence of the apparatus of its performance and the social dimensions that give rise to variation and 'error.' The proper object of linguistic study is therefore not speech itself but the grammar or knowledge that enables the performance of language. By privileging *competence* generative theorists choose to ignore the diversity of possible *performances,* and in doing so they idealize grammars as impervious to external factors. This opposition, then, denies a place for studies of language usage as ostensibly irrelevant to the study of I-language. But the exclusion of performance data leaves the discipline "on a dangerously unempirical base."[95] Data that contradict a theory of competence may easily be dismissed as irrelevant or anomalous to an I-language hypothesis. Both Chomsky and Saussure "make it clear that the real business of linguistic theory is

---

[94] Gregory Guy, "Competence, performance, and the generative grammar of variation," in *Variation, change and phonological theory*, ed. by Frans Hinskens, Roeland van Hout, and W. Leo Wetzels, Amsterdam Studies in the theory and history of linguistic science 146. (Philadelphia: John Benjamins, 1997): 125.
[95] Ibid.

langue/competence, and that the relationship between that theory and what happens when people use language is not something that their enterprise need be concerned with."[96] What data, then, are both relevant and critical to the study of music and mind, and what methods should be used in its collection and analysis? I propose that data should be both *relevant* in the sense that they are the product of specifically musical behaviors and processing and *critical* in the sense that basic musical operations cannot be conducted without them. The evidence of musical performances discussed in Chapter 4 shows that the study of pitch contours benefits from attention to the details of performances. If we exclude phonetic details of articulation, for instance, and rely entirely on preconceived categories of discrete pitches instead, then we ignore pitch elements that define the perceptual categories under investigation. 'Performance' data are thus crucial to establishing an empirical foundation for theories of music and mind.

Another limitation of research programs that make the competence/performance distinction is the assumption that it is sufficient to derive the innate structure of all languages (universal grammar) from a single one. Consider Chomsky's statement:

> I have not hesitated to propose a general principle of linguistic structure on the basis of observation of a single language. The inference is legitimate, on the assumption that humans are

[96] Ibid. Charles Hockett was also skeptical of this distinction. "We see the sense in which Chomsky is right in asserting that the relevant raw-material for a formal theory of a language is not what people say but what they *intend* to say. Nevertheless, what they actually do say is not to be ignored. Within the language system as characterized by a correct formal theory, all is determinate. But between the total workings of such a determinate system, or its analog inside a speaker, and the sound the speaker produces there is a layer of indeterminacy that can only be handled probabilistically." Charles F. Hockett, "Sound Change," *Language* 41, No. 2 (1965): 185-204. As a solution to this problem Hockett suggested models that are data-driven rather than theoretical (much like those of contemporary sociolinguistics and language usage).

> not specifically adapted to learn one rather than another human language. … Assuming that the genetically determined language faculty is a common human possession, we may conclude that a principle of language is universal if we are led to postulate it as a "precondition" for the acquisition of a single language.[97]

This assumption is made a priori and independently of comparative and cross-cultural investigations into a range of languages. Similarly for GTTM: Lerdahl and Jackendoff argue that an exclusive focus on Western art music is justified given that the underlying mechanisms are innate.[98] We need to be cautious in making such assumptions. As I pointed out in Chapter 1, there is no evidence supporting a biologically encoded language acquisition device, and the same goes for music. There is also considerable evidence to show that the extent of variation in human languages has been underestimated by generativists. For instance, Nicholas Evans and Stephen Levinson (2009) point out, "languages differ so fundamentally from one another at every level of description (sound, grammar, lexicon, meaning) that it is very hard to find any single structural property they share."[99] These manifestations do not undermine the hypothesis that there are shared cognitive mechanisms, but they do undermine the notion of a fixed set of rules and procedures, or 'knowledge of language.'

---

[97] Chomsky (1980, 48), quoted in: Nicholas Evans and Stephen C. Levinson, "The Myth of Language Universals: Language Diversity and its Importance for Cognitive Science," *Behavioral and Brain Sciences,* 32 (2009): 436.
[98] Lerdahl and Jackendoff, *A Generative Theory of Tonal Music*, 4.
[99] Evans and Levinson, "The Myth of Language Universals," 429.

Let us return briefly to the question of innateness so that the comparison between linguistic and music theoretical models is made clear. Ray Jackendoff frames the argument as follows: "[T]he central issue is what constitutes musical understanding, such that individuals can understand an unlimited number of pieces of music in a style with which they are experienced—and how through experience individuals acquire fluency in a musical style […]. Every normal individual has knowledge of language and music. Everyone learns the local variant(s) of both language and music. Normal adults are at ceiling for language, but they are more variable in musical ability, depending on exposure and talent."[100] The local variants that Jackendoff refers to (Chomsky's 'particular grammars') supposedly account for the diversity we find across cultures. The question is how much is learned, and how much is innate? "A formal theory of musical idioms will make possible substantive hypotheses about those aspects of musical understanding that are innate; the innate aspects will reveal themselves as 'universal' principles of musical grammar."[101]

> If the rules we have proposed correspond at all closely to principles unconsciously known and used by the experienced listener, one must ask how the listener manages to learn them. And of all the possible organizations one could attribute to tonal music (including all the incorrect ones posited by us music theorists), why does the listener infer the ones he does? The only answer that we find defensible is that one does not have to learn the entire grammar from scratch. Rather, one has no choice about much of it; many aspects of the grammar are simply the only (or easiest) ways

[100] Ray Jackendoff, 'Parallels and Nonparallels Between Language and Music," *Music Perception* 26, no. 3 (2009): 195.
[101] Lerdahl and Jackendoff, *A Generative Theory of Tonal* Music, 4.

> that one's mental abilities make available for organizing a musical signal. In other words, much of the complexity of musical intuition is not learned, but is given by the inherent organization of the mind, itself determined by the human genetic inheritance.[102]

These assumptions rule out explanations that account for many features of music and language, including usage-based models of statistical learning and pattern-recognition. The role of domain-general capacities is excluded from Jackendoff's account, and instead we are left with a restricted 'modular' analysis of how the mind/brain works. Jackendoff claims that, "musical intuition is not learned, but is given by the inherent organization of the mind, itself determined by the human genetic inheritance." But the biological and evolutionary basis for this proposal is largely conjectural. For Jackendoff and Chomsky's claims to be true both language and music must have evolved as specialized modules independently of one another and within the last one hundred thousand years. They would also have had to evolve by cognitive leap (mutation) rather than by incremental accumulation. This hypothesis is entirely opposed to principles of evolutionary theory and recent theories of human origins set out in Chapter 1. The generativist account also simplifies complex culture-historical processes to genetic determinants.

If the nature of musical grammar is disputed then so too are the elements upon which it is based. Rationalists and empiricists agree that music and language share combinatorial principles. However, it is not self-evident that these are composed of 'discrete elements.' Chomsky's view is dependent on a generative framework that assumes a stable set of

[102] Lerdahl and Jackendoff, *A Generative Theory of Tonal Music*, 281.

symbol tokens that do not vary and that do not change over time. Indeed, a tenet of generative theory is that a finite number of universal phonetic categories are used to generate an infinite number of utterances.[103] Chomsky and Halle proposed a theory of 'universal phonetics' to account for this limited set. And yet, despite the fundamental importance of this concept for generative grammar it remains at best a controversial one in linguistics: "Do phoneticians generally agree with phonologists that we will eventually arrive at a fixed inventory of possible human speech sounds? The answer is no."[104] Janet Pierrehumbert argues that, "[l]anguages differ systematically in arbitrarily fine phonetic detail. This means we do not want to think about universal phonetic categories, but rather about universal phonetic resources, which are organized and harnessed by the cognitive system."[105] Pierrehumbert conceives of this vowel space as a 'physical resource' that cultures 'divide up' according to their particular requirements. Unlike generative phonologists, phoneticians have tended to view language as a gradient rather than a discrete system. For instance, Ladefoged and Maddieson (1996) "assume that the space of possible speech sounds is indefinitely large." They are also explicit about the fact that

---

[103] "Although Descartes makes only scant reference to language in his writings, certain observations about the nature of language play a significant role in the formulation of his general point of view. In the course of his careful and intensive study of the limits of mechanical explanation, which carried him beyond physics to physiology and psychology, Descartes was able to convince himself that all aspects of animal behavior can be explained on the assumption that an animal is an automaton. In the course of this investigation, he developed an important and influential system of speculative physiology. But he arrived at the conclusion that man has unique abilities that cannot be accounted for on purely mechanistic grounds, although, to a very large extent, a mechanistic explanation can be provided for human bodily function and behavior. The essential difference between man and animal is exhibited most clearly by human language, in particular, by man's ability to form new statements which express new thoughts and which are appropriate to new situations" (Chomsky, *Cartesian Linguistics,* 59). The notion of discrete combinatoriality stems directly from these ideas and this inspiration.

[104] Robert F. Port and Adam P. Leary, "Against Formal Phonology," *Language* 81, No. 4 (2005): 927.

[105] Janet Pierrehumbert, "What people know about the sounds of language," *Linguistic Sciences* 29 (2000): 112.

"there is no closure to the set of possible speech sounds, thus finding themselves in sharp contrast with the phonetics of *Sound Pattern of English (SPE)* and most modern American phonologists."[106]

> Phoneticians in general, and at least some phonologists (Pierrehumbert 2000b, Bybee 2001), seem to deny the notion of an a priori, closed inventory of sound types in human language. Does this mean that categories of speech sounds are viewed as a mistake altogether? Are phoneticians simply denying the existence of any sound categories? No. Ladefoged and Maddieson and the IPA base their approach on the notion of distinct sound contrasts within languages. Certainly, the present authors take for granted that individual languages exhibit sound categories that yield contrasts between lexical entries. The many examples of minimal pairs show that languages do employ distinctive categories of speech sounds. *The difference between the generative school and the phoneticians has to do with the nature of the categories and how they are acquired and represented.* There is a critical difference between the notion of categories of sound and actual symbol tokens. The generative school insists that the sound categories are cognitive symbol tokens. But to call linguistic sound categories symbol tokens is a very strong assumption and one that need not be accepted in order to account for distinctive sound contrasts in languages [emphasis added].[107]

There is little evidence to support the notion of a mentalistic grammar of symbol tokens and it certainly does not correspond to what we know of linguistic diversity. Combinatorial systems appear to be universal but the notion that such systems are discrete and ordered according to strict rules and hierarchies is a far more controversial.

---

[106] Port and Leary, "Against Formal Phonology," 929.
[107] Ibid.

Do we really expect to find pitch categories that display similar properties (discrete, level pitch) and structure (functional hierarchies)?

To test this hypothesis let us consider the case of pitch in more detail. What would the relevant categories be? At this point it seems that pitch classes are the closest equivalent to the phonetic categories, or tokens, of Chomsky's universal phonetics. Studying tonal grammars and scale systems should help. Consider the following: Alexander Ellis long ago showed that scales and their structural properties are culture-specific constructions rather than innately determined properties of mind (1885).[108] I make a similar point in Chapter 4 where I show that even in Zulu music there is considerable diversity in scale systems. Not only that, but the pitch categories themselves are seldom 'discrete.' How are we to construct a universal phonetics out of this diversity? There appears to be no universal pitch class inventory across cultures. How pitch categories are structured in speech and song remains a complex theoretical problem. Pitch categories are culture-specific and syntactic features appear, in many cases, to be idiosyncratic, emerging from culture-specific conventions. What is shared across cultures then are not these details but rather the ability to acquire proficiency with one or other set of cultural variants. The only clear candidate for explaining such flexibility is domain-general learning mechanisms that integrate with culture-specific social learning. This perspective does not rule out combinatorial principles, but it does negate the idea of a universal phonetics and of discreteness as a categorical imperative for musical systems. As we pay more attention to

[108] Alexander Ellis, "On the musical scales of various nations," *Journal of the Society of the Arts* 33 (1885): 485-517.

the diversity within and across these systems in world cultures we need to develop a more flexible framework that we may term empiricist.

PART 3: EMPIRICISM

This dissertation puts forward an empiricist framework for song (and music) as the product of domain-general cognitive capacities and social learning. In Chapter 1 I introduced several of these concepts and laid the foundation by reviewing important work on acquisition and universals. These are important sources of evidence because they account for what seems to be a paradox: that music and language are universal but not innately specified.[109] This is because the similarities across languages reside in the mechanisms by which we perceive and produce utterances rather than in the categories used to construct them.

Functionalist approaches to language usage and cognition offer an alternative to the rationalist approach.[110] Joan Bybee (2010) and several others follow Joseph Greenberg in

---

[109] McClelland and Bybee argue that, "the evidence base for the existence of language universals is thin at best. Generative linguists tend to look for universals in one or two languages, since they believe languages are based on innate structures […]. More empirical approaches using samples of fifty, seventy-five or hundreds of languages provide a more valid basis for making claims about Universal Grammar. Yet researchers who do investigate large samples find many similarities among languages, especially in the way they change over time, but very few absolute synchronic universals of linguistic structure." James L. McClelland and Joan Bybee, "Gradience of Gradience: A reply to Jackendoff," *The Linguistic Review* 24: (2007), 443.

[110] Tomasello points out that, "hypothesizing the existence of an innate universal grammar brings with it two major acquisition problems that are currently unresolved—and that do not exist on a usage-based view. First is the problem of cross-linguistic diversity: How can the child link her abstract universal grammar to the particularities of the particular language she is learning (the linking problem)? Second is the problem of developmental change: How can we understand the changing nature of children's language across development if universal grammar is always the same (the problem of continuity)?" (Tomasello, *Constructing a Language*, 7).

demonstrating that there are no 'absolute' universals across languages.[111] Instead there are tendencies and patterns of change that are gradual and incremental. These alternative approaches are known as *usage-based linguistics* and demonstrate that languages possess *homologous* structural features despite their apparent diversity.

> Language is […] a phenomenon that exhibits apparent structure and regularity of patterning while at the same time showing considerable variation at all levels: languages differ from one another while still being patently shaped by the same principles; comparable constructions in different languages serve similar functions and are based on similar principles, yet differ from one another in specifiable ways; utterances within a language differ from one another while still exhibiting the same structural patterns; languages change over time, but in fairly regular ways. Thus it follows that a theory of language could reasonably be focused on the dynamic processes that create languages and give them both their structure and their variance. […] A focus on the dynamic processes that create language also allows us to move away from an exclusive focus on linguistic structures and formulate a broader goal: to derive linguistic structure from the application of domain-general processes. In this context, domain-general processes would be those that can be shown to operate in areas of human cognition other than language.[112]

[111] Greenberg tested linguistic universals across a wide range of languages and developed a linguistic typology. He concluded that most 'universals' are banal and tell us next to nothing about the nature of shared mechanisms. More recently, scholars such as Dryer (1998) have argued that, "statistical universals or strong tendencies are more interesting anyway. […] Where these tendencies are weak, they may reveal only bias in the current languages we have, or in the sampling methods employed. But where they are strong, they suggest that there is indeed a cognitive, communicative, or system-internal bias toward particular solutions evolving." Evans and Levinson, "The Myth of Language Universals," 438. See: Joseph H. Greenberg, ed., *Universals of* language (Cambridge, MA: MIT Press, 1963); Joseph H. Greenberg, *Language Universals* (The Hague: Mouton de Gruyter, 1966); Joseph H. Greenberg, "Diachrony, synchrony and language universals," in *Universals of Human Language: Vol. 1, Method and Theory*, ed. Joseph H. Greenberg, Charles A. Ferguson, & Edith A. Moravcsik, (Stanford, CA: Stanford University Press, 1978): 61-92.
[112] Joan Bybee, *Language, Usage and Cognition* (Cambridge: Cambridge University Press, 2010): 1.

By studying domain-general mechanisms we observe dynamic features of human cognition that are emergent rather than fixed or predetermined. That is, we see how the structure of language and music emerge from local, culture-specific practices with their own unique histories. Categories are acquired in richly textured learning environments (niches) that are best conceived as complex or open systems. In usage-based models, principles of *gradience* and *variation* are crucial to understanding how language and music operate as 'open' systems. "Gradience refers to the fact that many categories of language or grammar are difficult to distinguish, usually because change occurs over time in a gradual way, moving an element along a continuum from one category to another. […] Variation refers to the fact that the units and structures of language exhibit variation in synchronic use, usually along the continuous paths of change that create gradience."[113] These properties account for continuous change over time and are the basis for studying language and music as 'complex adaptive systems.'

> When linguistic structure is viewed as emergent from the repeated application of underlying processes, rather than given a priori or by design, then language can be seen as a complex adaptive system (Hopper 1987, Larsen-Freeman 1997, Ellis and Larsen-Freeman 2006). The primary reason for viewing language as a complex adaptive system, that is, as being more like sand dunes than like a planned structure, such as a building, is that language exhibits a great deal of variation and gradience.[114]

---

[113] Ibid, 2.
[114] Ibid.

Understanding music as a complex adaptive system is an attractive idea because, unlike the nativism of generative theory, it takes cultural variation to be an integral factor in the emergence of grammar. It gets us away from the inflexible innate 'competences' of rationalism and provides us with new conceptual tools for studying the performance of musics in real time (i.e. a processual model). If we accept that music is an open system in constant flux, then we see how its features are never completely stable and are always responsive to particular social pressures. Again, there is a clear parallel with language:

> Language has a fundamentally social function. Processes of human interaction along with domain-general cognitive processes shape the structure and knowledge of language. Recent research across a variety of disciplines in the cognitive sciences has demonstrated that patterns of use strongly affect how language is acquired, is structured, is organized in cognition, and changes over time. However, there is mounting evidence that processes of language acquisition, use, and change are not independent of one another but are facets of the same system. We argue that this system is best construed as a *complex adaptive system* (CAS). This system is radically different from the static system of grammatical principles characteristic of the widely held generativist approach. Instead, language as a CAS of dynamic usage and its experience involves the following key features: (a) The system consists of multiple agents (the speakers in the speech community) interacting with one another. (b) The system is adaptive; that is, speakers' behavior is based on their past interactions, and current and past interactions together feed forward into future behavior. (c) A speaker's behavior is the consequence of competing factors ranging from perceptual

> mechanics to social motivations. (d) The structures of language emerge from interrelated patterns of experience, social interaction, and cognitive processes.[115]

Music, too, has a "fundamentally social function" and it possesses many of these same emergent qualities, if in subtly different ways. Song emerges as a product of culture-specific patterns, and these patterns (conventions) are inherited and passed on through various linkages and lineages. This accounts for the fact that there is variation within and across cultures, at the global- (language-) level and the local- (dialect-) level. Practices change to meet new social and contextual functions, whether these are religious, political or interpersonal. Changes such as these have nothing to do with innate mechanisms. Any number of factors may be invoked so long as they are intelligible and appropriate to specific communities. That is, so long as they have *functional* necessity. It is the social dimension of human cognition that seems to have driven our cognitive evolution. It may well be that language and music are the products of mechanisms that evolved precisely for these forms of complex sociality.

Some of the best evidence for usage-based models has emerged in studies of language acquisition. In his summative work, *Constructing a Language* (2003), Michael Tomasello posited two main sets of learning mechanisms to account for language acquisition: intention reading (theory of mind) and pattern-finding (categorization). The first set of skills "are necessary for children to acquire the appropriate use of any and all linguistic

---

[115] Clay Beckner, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann, "Language is a Complex Adaptive System: Position Paper," *Language Learning* 59, Suppl. 1 (2009): 2.

symbols, including complex linguistic expressions and constructions. Indeed, they basically define the symbolic or functional dimension of linguistic communication—which involves in all cases the attempt of one person to manipulate the intentional or mental states of other persons."[116] The second set of mechanisms enable us to accomplish, inter alia, the following: to "form perceptual and conceptual categories of similar objects and events; form sensory-motor schemas from recurrent patterns of perception and action; perform statistically-based distributional analyses; and produce analogies."[117] All of these enable children to find patterns of usage from across a wide range of utterances, and to "construct the grammatical (abstract) dimension of human linguistic competence."[118] This grammatical dimension is critical because it undergirds the combinatorial features of both language and music (to be discussed in more detail in Chapters 3 and 4). Tomasello's mechanisms of "intention-reading" and "pattern-finding" may guide new studies into music usage and acquisition. In the following chapters I focus on pattern-finding mechanisms since these are important for understanding the structure of melodies across cultures, and I investigate the categories upon which these mechanisms operate.

An important theme in this study is the focus on non-discrete, gradient elements in the structure of song that I have termed prosodic. Accommodating these elements means attending to chunks of material of various shapes and forms, in other words, to categories

---

[116] Tomasello, *Constructing a* Language, 3.
[117] Ibid.
[118] Ibid.

that do not reduce to individual 'pitches.' The usage-based model provides theoretical support for this focus because it emphasizes the fact that elements are encoded based on their functional relevance in contexts of performance and communication. Melodies may thus consist of a variety of different categories that are culture-specific, even if the mechanisms by which they are structured are the same. We see this parallel in usage-based linguistics as follows: "competence with a natural language consists of the mastery of all its items and structures, and these constitute a much more complex and diverse set of linguistic representations than the 'core grammar' of formal approaches."[119] Tomasello's work reinforces this argument because he shows that the units of language that children acquire include small and large structures—often whole linguistic utterances—and that these have *functions* (i.e. they are not simply abstract elements in a combinatorial system).

In sum, the usage-based approach has three main features: First, it is *functionalist* in the sense that it is "based explicitly in the expression and comprehension of communicative intentions" in a particular culture. Second, it is *construction-based*: the focus is not on 'discrete' elements or pitch-categories alone, but rather on the variety of elements that are communicatively salient. These may include individual pitch categories as well as larger chunks of material. Third, it is *usage-based*: the grammaticalization of song emerges from use—both historically and ontogenetically. Taking these three principles into

[119] Ibid, 5-6.

account, I outline a theory of melody as prosody in Chapter 3, and provide empirical support for this with a case study of Zulu song prosody in Chapter 4.

Conclusion

This chapter has set out a broad critique of rationalist approaches to language, music, and mind. Generativists, in particular, make strong assumptions about the operation of human cognitive processes (mentalism), the biological basis for language and music (nativism), and how these behaviors are structured in the brain (modularity). But, as I pointed out in Chapter 1, these assumptions are increasingly tenuous given recent evidence from studies of language acquisition, brain localization, and universals. They are also not consistent with accumulating evidence for linguistic and musical diversity in world cultures. Generativists seem to overlook the importance of domain-general cognitive capacities by giving priority to what argue are domain-specific adaptations for language or music. Rationalists also rely almost entirely on indirect evidence of 'competence' to substantiate their claims rather than on direct evidence of actual performances. By eliminating performance from studies of music and mind they tend to ignore perceptual cues and categories that are crucial to our perception and cognition of language and music.

Empiricists, on the other hand, propose that grammar emerges through usage as a product of domain-general cognitive capacities and social learning. Usage-based theory takes into account the variation and gradience we find across cultures and provides a model for studying song as a complex adaptive system. The case study for this dissertation puts this program into practice by studying evidence of actual performances rather than musical scores, and by focusing on the complexity of the interrelated linguistic and musical phenomena that emerge in Zulu song. The categories of song prosody described in later chapters are understood to be the products of culture-specific learned practices rather than the innate rules and procedures that rationalist claim to be universal. An empiricist account of music, language, and mind is consistent with evidence from across the full spectrum of the cognitive sciences because it builds on this evidence rather than logic alone.

CHAPTER 3

Melody as Prosody

Melody is conventionally defined as a set of discrete pitches patterned in time.[120] But pitch patterning in song includes gradient features that cannot be explained by models of discrete pitch alone. This chapter describes the range of pitch categories characteristic of song melody and how they are combined into a multi-dimensional framework, what in the previous chapter I described as a 'complex adaptive system.' I survey conventional explanations for pitch phenomena including the music-specific elements of tonality, discrete pitch, hierarchical structure, and contour. In keeping with the usage-based framework developed in the previous chapter, however, I supplement these with several other concepts, including: pitch targets, glides and chunking. To these features and processes are added the prosodic elements of tone and intonation. Together they account for the gradience and variation we find in melodies across cultures. Taking all of these factors into consideration, I outline a model of prosodic structure in which the syntagmatic and paradigmatic elements of pitch patterning in song are complexly ordered into a multi-dimensional, networked pitch space. Because these various elements are not

[120] Definition of melody: "In the most general sense, a coherent succession of pitches. Here pitch means a stretch of sound whose frequency is clear and stable enough to be heard as not noise; succession means that several pitches occur; and coherent means that the succession of pitches is accepted as belonging together." Don Randel, ed. "Melody [fr. Lat. Melodia, Fr. Gr. Meloidia, Fr. Melos]." *The Harvard Dictionary of Music*. Cambridge: Harvard UP, 2003. *Credo Reference*. 25 June 2004. Web. 5 Nov. 2013. <http://0-www.credoreference.com.oasis.unisa.ac.za/entry/harvdictmusic/melody_fr_lat_melodia_fr_gr_meloidia_fr_melos>.

neatly divisible into discrete categories I treat melody *as* prosody; an approach that I apply to a case study of Zulu song prosody in Chapter 4.

PART 1: MELODY

Current research on melody suffers three biases that were addressed in the two previous chapters, including an almost exclusive focus on (1) instrumental music; (2) Western art music of the common practice era; and (3) the musical score as constitutive of musical experience. These biases are evident in Aniruddh Patel's claim that "a musical melody is an aesthetic object, a sound sequence that is an end in itself, whereas a linguistic intonation contour is simply a means to an end, in other words, pitch in the service of quotidian linguistic functions."[121] This aestheticizing bias is characteristic of formalist approaches in which music's 'ephemeral' qualities are partitioned off as supplementary and its structural and syntactic qualities given priority. A goal of this study is to contest the notion of melody as an end in itself. Melody as defined here is not equivalent to *a* melody, that is, a discrete pitch array independent of expressive purpose, intent, or illocutionary force. As communicative acts, song melodies are meaningful beyond the syntactic structure of discrete pitches. Humans are calibrated to perceive microscopic changes in pitch more nuanced than pentatonic, heptatonic, or chromatic scale divisions and their intervallic structures suggest. Gradient phonetic cues play a crucial role in our perception and cognition of song. The function of such cues becomes clear when we investigate the prosodic structure of pitch relations rather than their geometric division in

---

[121] Aniruddh Patel, *Music, Language, and the Brain* (New York: Oxford University Press, 2008): 184.

pitch space. Pitch categories function within several layers or levels of patterning that are both local and global. Local events are organized into a linear structure of oppositional relations, but there are also combinatorial and grammatical principles involved in the organization of local events that are global. How these fit with hierarchical cognitive processes in a network of relationships is a challenge for cognitive music theory. What this chapter aims to show, at the very least, is that the form and function of pitch patterning in song has a structure and function beyond that 'determined' by composers. I treat song as a flexible, robust, and natural communicative system that is usage-based rather than innately acquired, and I provide a processual model for its study.

This approach to the study of melody opens a dialogue with several cognitive music theorists. It shares commonalities with the 'Penn School' of music theory.[122] Leonard Meyer and Eugene Narmour, each in his own way, provide a starting point for modeling melody based on cognitive principles, and many others have applied experimental methods to test the hypotheses presented in their work.[123] The Penn school developed a

---

[122] Robert Gjerdingen first used the term 'Penn School of Music Theory' in his article, "The Psychology of music," in *The Cambridge History of Western Music Theory*, ed. Thomas Christensen (Cambridge: Cambridge University Press, 2008). The approach advocated by Meyer and then Narmour was taken up by several of their students including Gjerdingen himself, in his work on schema, and by Justin London and Chris Hasty in their work on rhythm and meter respectively. See: Robert Gjerdingen, *A Classic Turn of Phrase: Music and the Psychology of Convention* (Philadelphia: University of Pennsylvania Press, 2008); Justin London, *Hearing in Time: Psychological Aspects of Musical Meter* (New York: Oxford, 2004); Christopher Hasty, *Meter as Rhythm* (Oxford: Oxford University Press, 1997).

[123] See: Timothy C. Justus, Jamshed J. Bharucha, "Music Perception and Cognition," in *Stevens' Handbook of Experimental Psychology*, Third Edition, ed. H. Pashler, H. and S. Yantis (New York: Wiley, 2002); Carol Krumhansl, *Cognitive Foundations of Musical Pitch*, Second Edition (New York: Oxford, 2001); Mark A. Schmuckler, "Pitch and Pitch Structures," in *Ecological Psychoacoustics*, ed. J. Neuhoff (San Diego, CA: Academic Press, 2004): 271-315; Allan Vurma and Jaan Ross, "Priorities in Voice Training: Carrying Power or Tone Quality," *Musicae Scientiae* 4 (Spring 2000): 75-93; Mark A. Schmuckler and

cognitive music theory grounded on psychological principles that foreground the *listening experience*. Unlike the work of generative theorists who model not the experience of music but the 'competence' of an idealized listener, the Penn School factors processual elements of music perception and cognition.

In Meyer's foundational 1956 monograph, *Emotion and Meaning in Music*, he outlined an influential vision for the study of music as a domain of psychology.[124] His interdisciplinary focus grounded scholars in the new discipline of music cognition, and some of his ideas led to what became a consensus in the field: that idea that *expectation* is key to the experience of emotion and meaning in music. This thesis is embodied in a statement Meyer makes early on in his important book:

> Affect or emotion-felt is aroused when an expectation – a tendency to respond – activated by the musical stimulus situation, is temporarily inhibited or permanently blocked […]. Embodied musical meaning is, in short, a product of expectation. If, on the basis of past experience, a present stimulus leads us to expect a more or less definite consequent musical event, then that stimulus has meaning.[125]

Robert Tomovski, "Perceptual Tests of an Algorithm for Musical Key-Finding," *Journal of Experimental Psychology: Human Perception and Performance* 31, No. 5 (2005): 1124-1149.

[124] Leonard B. Meyer, *Emotion and Meaning in Music* (Chicago: University of Chicago Press, 1956); See also: Grosvenor Cooper and Leonard B. Meyer, *The Rhythmic Structure of Music* (Chicago: University of Chicago Press, 1960); Leonard B. Meyer, *Explaining Music: Essays and Explanations* (Chicago: University of Chicago Press, 1973).

[125] Meyer, *Emotion and Meaning in Music*.

Meyer places experience at the very center of his theory. For expectation cannot be isolated, it must be the product of past and present experience combined. He argued that expectations were 'grammatical' to experienced listeners of tonal music. Unlike generative theorists, who also study 'experienced listeners' but for whom innate principles guide their perception of musical structure, Meyer's work emphasizes the role of learning. "It is possible that the laws of the mind may in some circumstances be independent of cultural conditioning. Where human communication is involved, however, though the laws still operate, they do so within a socio-cultural context where attitude, belief, and learning qualify their operation."[126] Meyer recognized what we may term domain-general mechanisms that underscore pattern recognition, including the Gestalt principles of proximity, similarity and common fate.

> The work of the Gestalt psychologists has shown beyond a doubt that understanding is not a matter of perceiving single stimuli, or simple sound combinations in isolation, but is rather a matter of grouping stimuli into patterns and relating these patterns to one another. And finally, the studies of comparative musicologists, bringing to our attention the music of other cultures, have made us increasingly aware that the particular organization developed in Western music is not universal, natural, or God-given.[127]

Learning builds on these principles: "The vital role occupied by learning in conditioning the operation of Gestalt laws and concepts indicates at the outset that any generalized Gestalt account of musical perception is out of the question. Each style system and style

---

[126] Meyer, *Emotion and Meaning in Music*, 84.
[127] Meyer, *Emotion and Meaning in* Music, 6.

will form figures in a different way, depending upon the melodic materials drawn upon, their interrelationships, the norms of rhythmic organization, the attitudes toward texture, and so forth."[128] In other words, within a circumscribed cultural practice or style there are learned conventions that condition expectation. Composers internalize these conventions and manipulate them to generate melodic interest and affect.

In sum, music perception and cognition is the product of a system of tonal relations that are learned both intuitively through development and consciously through education. But music cognition is not entirely learned. In cognitive music theory there are still many who studied isolated stimuli as the basis for theorizing melody. By invoking Gestalt principles Meyer was able to point to holistic features of music perception that are too often ignored by those who study such isolated phenomena. He showed that meaning in music is not entirely arbitrary (the product of social construction) and that it is the product of basic responses to sound that are governed by domain-general cognitive capacities. These are crucial to the concept of 'tonality.'

3.1 Tonality

A system of tonality is dependent on an implicit ordering of pitches that create a sense of tension and repose based on their relative distance from a central tonic. A tonic is generally conceived as a single tone, chord, or pitch center. This tonic acts as an attractor

---

[128] Meyer, *Emotion and Meaning in* Music, 85.

to other tones or chords in a pitch space and sets up a hierarchy of relations that are used to generate melodic structure.[129]

> The term “tonality” refers to the relationships existing between tones or tonal spheres within the context of a particular style system. As Strangways has put it: “A tonic is a tendency rather than a fact.” That is, some of the tones of the system are active. They tend to move toward the more stable points in the system—the structural or substantive tones. […] But activity and rest are relative terms because tonal systems are generally hierarchical: tones which are active tendency tones on one level may be focal substantive tones on another level and vice versa. […] At the other end of the architectonic scale it should be noted that tonality plays a part in the articulation of musical forms larger than those of the phrase or melody.[130]

Tones move toward points of repose, or tonics, and even ‘non-essential’ tones in a melody or scale play a vital role in establishing tonal centers, and these give rise to an emergent hierarchical structure. The features of tonality are learned through repetition within particular cultural contexts, but for Meyer they also rely on Gestalt principles. Listeners learn tonal conventions both in the moment and through consistent conditioning. Narmour codified these insights by saying that listeners possess a sense of *intraopus* style (conventions established within the work itself) and *extraopus* style (conventions of a larger repertory stored in memory through prior learning). In other words, listeners’ perceptions may be altered and shaped through immediate tonal

[129] The term ‘attractor’ is most often used in dynamical systems theory to describe a point or set of points to which a system evolves. This is consistent with the idea of a pitch target (a concept developed later in this chapter) to which a singer orients a pitch contour or glide within a larger pattern. The benefit of using the term ‘attractor’ is that its target need not be defined by a discrete point in a pitch space (that is, a level or absolute pitch). Rather, this cognitive target may have gradient qualities.
[130] Meyer, *Emotion and Meaning in* Music, 214-215.

immersion in a particular song but also through long-term exposure to similar types of tonal organization in many songs.

The fact of tonality is undisputed in all sectors of cognitive music theory and is an important scaffold upon which to build a theory of melody. Jackendoff and Lerdahl argue that a tonic attractor is the crucial and defining element.

> What is representative of the world's musical idioms is not harmony, but rather a broad sense of tonality that does not require or even imply harmonic progression and that need not be based on the familiar Western major and minor scales (Nettl, 1973). Western harmony is a particular cultural elaboration of this basic sense of tonality. In a tonal system in this sense […], every note of the music is heard in relation to a particular fixed pitch, the *tonic* or *tonal center*. The tonic may be sounded continuously throughout a piece, for instance by a bagpipe drone or the tamboura drone in Indian raga; or the tonic may be implicit. Whether explicit or implicit, the tonic is felt as the focus of pitch stability in the piece, and melodies typically end on it. Sometimes, as in modulation in Western music, the tonic may change in the course of a piece, and a local tonic may be heard in relation to an overall tonic. The presence of a tonal center eases processing (Deutsch, 1999) and is a musical manifestation of the general psychological principle of a cognitive reference point within a category (Rosch, 1975).[131]

According to Jackendoff and Lerdahl, this tonal system operates in a pitch space situated in relation to the tonic. Other pitches (the musical scale) are arranged in fixed intervals relative to the tonic and are what enable hierarchical structure. "On both empirical and

---

[131] Jackendoff and Lerdahl, "The Capacity for Music," 45.

theoretical grounds (Krumhansl, 1990, TPS, chapter 2), cognitive pitch-space distances are hierarchically organized. There is even provisional evidence that these structures have brain correlates (Janata et al., 2003)."[132] It remains to be seen whether this pitch space model and its hierarchical and fixed intervallic components are characteristic of all tonal systems across cultures. This seems to be a strong assumption, given the wide range of traditions in which clearly articulated modal and functional harmonic frameworks are not operative. Not all cultures designate scale degrees with clearly defined intervallic ratios.

The idea that tonality operates as an orientation device with a differentiated pitch space has strong empirical support in music cognition. Patel writes that, "pitches in musical melodies are organized into scales that feature salient hierarchies of stability. These hierarchies are reflected in the systematic use of different scale degrees for different structural purposes (e.g., scale degree 1 as a stable resting point). As a result, different scale degrees take on distinct psychological qualia in the fabric of the music."[133] For Meyer, tonality is the 'binding force' that gives meaning to musical style: "One of the most powerful and persuasive forces conditioning and controlling the sense of completeness which a melodic pattern gives is the tonal organization or scale of the culture. It establishes a system of order, a system of expectancies, a system of tonality […] Tonality is probably the most important single facet of style, the *sine qua non* of even the most primitive musical organization."[134] Yet the near-universal status accorded

---

[132] Jackendoff and Lerdahl, "The Capacity for Music," 46.
[133] Patel, *Music, Language, and the Brain*, 201.
[134] Meyer, *Emotion and Meaning in* Music, 138.

scale degrees and the functional hierarchical relations of common practice tonality are not consistent with musics in many world cultures.[135] For as Krumhansl and Kessler put it, "The perception of musical structure depends on the processing of pitch information with reference to a system of knowledge about the conventional uses of pitches within the musical tradition."[136] These "conventional uses of pitch" are crucial to learning the particular combinatorial system of a music culture. But I will argue that scale degrees and pitch class relations are not the only or even the primary elements to melodic structure in song cultures.

An important part of this argument about tonality is the need to avoid 'totalizing' systems. In *Beyond Schenkerism* (1977) Eugene Narmour signaled dissatisfaction with theories of tonality that treat music as a 'whole system':

> The major tenet (and the major problem) of transformational grammar and Schenkerian theory is the notion that language, musical or otherwise, is a *whole system*. This leads to the belief that linguistic usages at any given time can be described independently of particular statements. In Schenkerian theory, for instance, neighboring-tone structures, passing-tone structures, and so

---

[135] "It is important also to recall that the norms of style, of tonality, important as they are, can be altered through training and knowledge. […] [Hence], what we know, either because we have been told or have learned through practice and experience, influences our judgment of what we perceive and hence our feeling of completeness and our expectation based upon that feeling" (Meyer, *Emotion and Meaning in Music*, 138).

[136] Carol Krumhansl and Edward J. Kessler, "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys," *Psychological Review* 89 (1982): 334-368.

> forth, have independent status as transformational structures and are believed to be governed in some sense by rules which are not dependent on context.[137]

Narmour argues that formalist theories of this sort have no basis in perception. In other words, we need to pay more attention to *usage*. This approach emphasizes the 'diachronic implications' of melody: "[L]ike transformational grammar, Schenkerian theory, with its *Ursatz*-kernel and its assumption of tonal language as a whole system without diachronic implications, accounts for the syntactic structure of a work by idealizing pitch relationships and disembodying them from much of their relevant data—for example, their rhythmic, metric, formal, melodic, and, in some cases, registral characteristics."[138] Narmour also cautioned against approaches that "[do] not distinguish compositional structures from perceptual structures,"[139] a position very much in keeping with the usage-based approach developed in this study.

Narmour's own theory of melody is the implication-realization (I-R) model, first introduced in *Beyond Schenkerism* but brought to fruition in two major monographs: *The Analysis and Cognition of Basic Melodic Structures* and *The Analysis and Cognition of Melodic Complexity*.[140] In these works Narmour examined the note-to-note relations of

---

[137] Eugene Narmour, *Beyond Schenkerism: The Need for Alternatives in Music Analysis* (Chicago: University of Chicago Press, 1977): 167.
[138] Narmour, *Beyond* Schenkerism, 171.
[139] Eugene Narmour, "Some Major Theoretical Problems Concerning the Concept of Hierarchy in the Analysis of Tonal Music," *Music Perception* 1, no. 2 (1983): 129.
[140] Eugene Narmour, *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model* (Chicago: Chicago University Press, 1990); Eugene Narmour, *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*, (Chicago: University of Chicago Press, 1992).

melody to establish 'rules,' or what in the spirit of scientific inquiry he termed 'hypotheses.' He developed a complex parametric toolbox for unpicking the 'genetic code' of melody. Drawing on the Gestalt principles of similarity, proximity, and common fate that Meyer introduced in his important book, Narmour claimed that both bottom-up and top-down psychological processes act simultaneously in the cognition of melody.

> The implication-realization model hypothesizes that emotional syntax in music is a product of two expectation systems—one top down, the other bottom up. Syntactic mismatch or conflict in realizations can occur either within each system or between them. The theory argues that interruption or suppression of parametric expectations generated separately by the two systems explains certain types of recurrent aesthetic strategies in melodic composition and accounts for the most common kinds of musical forms (AAA, AAB, ABB, ABC, and ABA).[141]

The two systems include both learnt stylistic tendencies and hard-wired psychological impulses that are impervious to such learning.

> The top-down system is flexible, variable, and empirically driven. In it, listeners constructively match and compare representative schemata to current input. Schemata range from highly instantiated parametric complexes within a style to extremely generalized structurings of the elementary materials of a style (e.g., scale-step hierarchies in tonal music). Musically, this top-down system divides into intra- and extraopus style, where both prior learning before listening to a piece and immediate learning during a piece influence expectation. […] In contrast, the bottom-up mode constitutes an automatic, unconscious, preprogrammed, 'brute' system that operates on

[141] Narmour, *The Analysis and Cognition of Basic Melodic Structures*, 3.

> parametric primitives (e.g. intervals, registral directions, durations, perceptual consonances and dissonances).[142]

The two systems provide a comprehensive framework for unpicking the structure of melody and the different sorts of cognitive factors crucial to its cognition. In many ways Narmour's work builds on and codifies Meyer's insights. But Narmour goes much further in establishing a rigorous objective framework for our investigations. Not only does he bring his theory in line with findings in cognitive psychology and neuroscience, but he also minimizes the role of the analyst as arbiter; the traditional emphasis in music theory on intuitive inferences is largely expunged from this model with Narmour warning against introspection as a mode of inquiry. [143] The I-R model also goes much further than Meyer in specifying which processes are learned through experience and which are 'hard-wired.' He insists that bottom-up processes are at the very root of melody. These 'parametric primitives' are not necessarily discrete entities and are rather like 'rules of continuation'. The I-R theory provides a functional explanation for how the musical mind works and how the processes of melodic syntax are understood.

---

[142] Narmour, *The Analysis and Cognition of Basic Melodic Structures*, 4.

[143] In her review of *The Analysis and Cognition of Basic Melodic Structures*, Naomi Cumming felt alienated by the measure of objectivity Narmour imposes on analysts: "Narmour makes music theory a sub-discipline of cognitive psychology. I find accordingly that his theory excludes me as a subject listening to music, and replaces my 'I' with a cognizing brain. The theory is concerned not with conscious experience, but with the automatic procedures of information-processing which are neuronally 'hard-wired'. A theorist who clings to the value of conscious musical experience within a learned culture as the context in which analytical decisions are made might well resist Narmour's treatment of melodic structures as bits of information automatically processed in the brain, but the search for cognitive universals is not, in itself, reprehensible, and rejection of his theory cannot be justified on this ground" (Naomi Cumming, "Eugene Narmour's Theory of Melody," *Music Analysis* 11, Nos. 2-3 (1992): 355). In Chapter 2 I was critical of approaches to cognitive music theory that rely on intuition as their guide, including the work of Meyer, and Lerdahl and Jackendoff in GTTM. Narmour's approach is different in that it emphasizes bottom-up processing that is 'hard-wired.' More importantly, it imparts a measure of 'objectivity' that is necessary if we are to move beyond subjective assumptions that may or may not be grounded in cognitive reality.

What is useful to this study are not so much the basic structures and processes that Narmour uses in his analysis—since this study is not directed at explicating syntactic structures and their meanings—but rather the way in which Narmour approaches the study of melody as a complex adaptive system that is not bound to the conventions of tonality assumed by most theorists. "What makes Narmour's approach unique is his temporary evasion of both tonality and style in an attempt to explain perceptual responses from the 'bottom-up', dealing only with those cognitive components which may be construed as pan-stylistic or universal."[144] Another reason why I have not tested the specific structures of the I-R model is because they are best suited to studies of notated Western tonal music. This is not to say that the cognitive processes underlying this model are culture-specific—Narmour explicitly claims that they are not—but rather that my study is explicitly concerned with problematizing the categories of melodic structure assumed in its implementation to tonal music. The purpose of this chapter is to theorize the gradience and variation indicative of prosodic-melodic categories in vocal music. It is important, therefore, that we investigate the role of discrete pitch and hierarchical structures predicated on scale degree.

3.2 Discrete pitch

The orthodoxy of cognitive music theory today, as I pointed out in section 3.1, is that melodies consist of discrete intervallic structures and stable tonal relations. "One of the striking commonalities of musical pitch systems around the world is the organization of

---

[144] Cumming, "Eugene Narmour's Theory of Melody," 355-356.

pitch contrasts in terms of a musical scale, in other words, a set of distinct pitches and intervals within the octave that serve as reference points in the creation of musical patterns."[145] This consensus extends to what is proposed as a key difference between music and language: that musical melody consists of discrete pitch events whereas linguistic melody is *continuously varying*. But this distinction is made on the basis of instrumental music, not song. I will show that such gradient elements are in fact distinctive of song melody. The problem with discrete categories, as employed by generativists, is that they are rigid and absolute—they are either present or absent and cannot be partial. To base our analysis on these categories would ignore a range of prosodic pitch phenomena and so I introduce several alternative categories instead. In place of discrete pitch, for instance, I propose the concept of *pitch targets*. This enables us to include many features of discrete pitch without reifying discrete tonal identities as absolute.

The idea of discrete pitch emerges from a digital model of cognition in which the mind combines discrete conceptual units into a potentially unlimited number of combinations to produce coherent meanings. In generative theory there are rules and constraints that govern these combinations (see Chapter 2). Discrete pitch categories are assumed to be the fundamental units in these computationist models. Consider Jackendoff's statement: "Unlike any other cognitive capacities, both language and music involve a sequence of

---

[145] Patel, *Music, Language, and the Brain*, 14.

digitized sounds: speech sounds in language, tones or pitch events in music."[146] But the uses of pitch in music and language differ: "only melodies have discrete pitches, while prosodic contours usually involve a continuous rise and fall."[147] Patel makes a similar point: "The lack of a stable interval structure is the single most salient difference between musical and spoken melodies, and is likely why theories of musical melody and theories of speech melody have had so little conceptual contact."[148] Again, these theories of discreteness are convenient in making a firm distinction between music and language, but they simply ignore features of gradience and variation in pitch organization. After all, the categories of perception need not be discrete. "[N]onmusicians [do] not show strong evidence of categorical perception. This evidence strongly suggests that intervallic relations and their perception is the product of learning rather than innateness."[149] Diverse pitch categories are possible and indeed probable.

There is cross-cultural evidence of gradience and variation in pitch patterning. J.H. Kwabena Nketia has elaborated this in his studies of African music:

> The scales used in vocal music, having from four to seven steps, are not unlike those used in instrumental music. For any given scale step, however, one may not always find absolute correspondence between the vocal pitches and the instrumental tunings. In African musical

[146] Jackendoff, "Parallels and Non-Parallels Between Language and Music," 198.
[147] Jackendoff, "Parallels and Non-Parallels Between Language and Music," 199.
[148] Patel, *Music, Language, and the* Brain, 206.
[149] Patel, *Music, Language, and the Brain*, 25.

> practice, the areas of tolerance of pitch variation for particular steps of the scale are much larger than those of traditions that base their music on a fixed pitch of 440 vibrations per second for A.[150]

A cognitive explanation is required for this "tolerance." If pitch units exhibit a high degree of gradience then there must be a significant degree of perceptual adjustment, a shift in expectations. Were discrete pitch in operation we would expect there to be narrowly defined thresholds beyond which new, distinct categories of pitch units are formed. If we take pitch categories to be more fluid, then this "tolerance" for pitch variation is accounted for by a general cognitive adjustment for a gradual narrowing of tessitura over the course of the utterance. This is a phenomenon phoneticians describe as 'pitch-span reduction' and it is common to most speech utterances. There is no reason to believe that it does not operate similarly in song. In other words, the capacity to recognize distorted or reduced pitch contours as possessing the same pitch identity is a crucial feature of music cognition.

Still, the concept of 'discrete pitch' has played a vital role in accounting for features of tonality and hierarchical structure for many music cultures, and any new model must account for these functional attributes. Patel reviews the important of intervallic structures to musical melody and makes a comparison with tonal linguistic features as follows:

---

[150] J.H. Kwabena Nketia, *The Music of Africa* (New York: Norton, 1974): 147.

> [A] stable system of intervals allows musical melodies to make use of a tonal center, a focal pitch that serves as a perceptual center of gravity for the melody. An interval system also allows the creation of a hierarchy of pitch stability in melodies. In contrast, the 'tones' of intonation have no such organization: Each tone is used where it is linguistically appropriate and there is no sense in which some are more stable or central than others. Another consequence of an interval system is that when combined with a temporal grid provided by beat and meter, a scaffolding is created for an elaborate set of structural relations between tones. This is likely part of what makes musical melodies so aesthetically potent. In contrast, the network of pitch relations in intonation contours is not nearly as rich. As a result, intonation contours are aesthetically inert, as evidenced by the fact that people rarely hum intonation contours or find themselves captivated by the pitch patterns of speech. This is quite sensible, as a musical melody is an aesthetic object, a sound sequence that is an end in itself, whereas a linguistic intonation contour is simply a means to an end, in other words, pitch in the service of quotidian linguistic functions.[151]

Apart from assumptions about the respective functions of music and language, Patel's comparison of melody and intonation is problematic in several respects: for one thing, intonation cannot be segregated from other pitch elements in either song or speech. Second, non-instrumental melodies are far more graded than his analysis admits. And even more important is the fact that many linguistic factors impose special constraints on musical melody too. The focus on discrete pitch is problematic in each of these respects and needs to be supplemented. We need to account for the many important elements of melodic structure associated with pitch events but we need to do so while at the same time ensuring that gradient elements are not ignored. This means devising a model of

[151] Patel, *Music, Language, and the Brain*, 183-184.

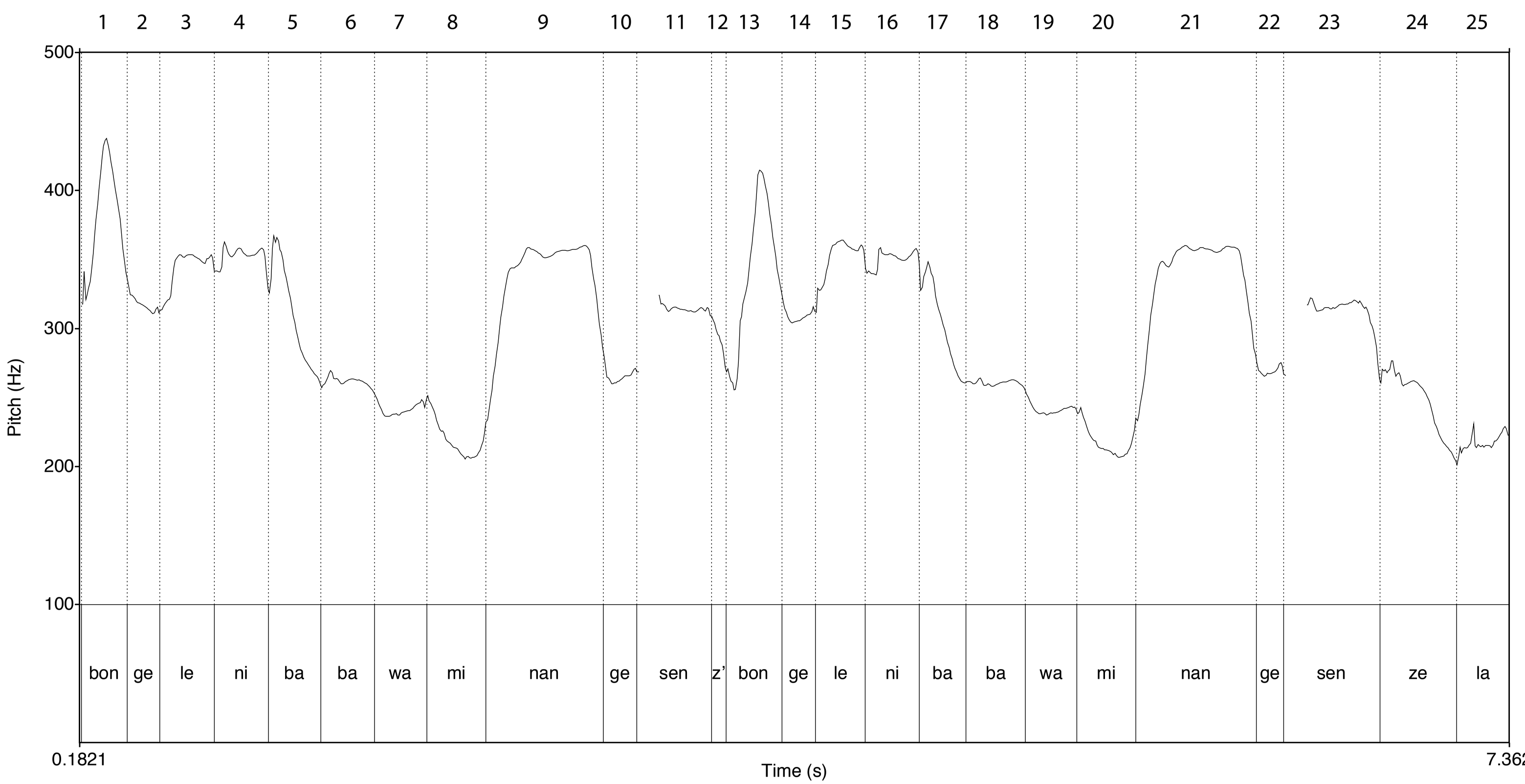
melodic structure that factors both sorts of categories. To accomplish this task I introduce the notion of pitch targets, thereby retaining the element of combinatorial organization in melody without assuming a rigid structure of discrete elements.

3.3 Pitch Targets

A *pitch target* is a focal point to which there are various forms of ascent or descent. The consistency of pitch targets allows for cognitive anchoring and consistency of pitch height and tessitura in a melody, and in some songs may be used to establish a tonal center. Pitch targets do not require level or absolute pitch realizations for their efficacy. Pitch targets of this sort are clearly in operation in most musics, since combinatorial mechanisms are dependent on such categories to allow for tonal organization of even the most rudimentary kind. Pitch targets are culture-specific and mostly learned. However, there are certain proclivities of the human auditory apparatus that include the ability for perceiving octave equivalence and a predilection for the fifth. Still, apart from this rather broad segmentation of pitch space, there are a wide variety of possible scales and configurations. The following example of a Zulu song demonstrates the value of pitch targets as categories in song.

Example 1 shows a pitch track (fundamental frequency reduction) of the Zulu *memulo* song *Baba Wami* (for a more detailed analysis and explanation of the genre please see Chapter 4, Section 3). Pitch is measured in hertz on the Y-axis with time on the X-axis. The words of the lyric are divided into segments that correspond to pitch events. These

Example 1: *Baba Wami* pitch track showing features of gradience



Example 1 shows a pitch track (fundamental frequency reduction) of the Zulu *memulo* song *Baba Wami*. Pitch is measured in hertz on the Y-axis with time on the X-axis. The words of the lyric are divided into segments that correspond to pitch events. These are numbered at the top of the graph. The example may be divided into two halves (1-12; 13-25) that repeat the same phrase. There are consistent pitch targets even though there are significant differences in their precise realization in the two phrases. For instance, segment 1 is noticeably higher in frequency than segment 13 even though they appear as the first segment in both phrases and perform the same function (a high tone followed by a gradual descending contour). There are few moments of discrete pitch and many instances where there is an apparent peak (1), trough (7) or overshoot (9). Glides are also evident at (5), (17) and (24).

are numbered at the top of the graph for ease of use. The example may be divided into two halves (1-12; 13-25) that repeat the same phrase. There are consistent pitch targets through the course of the song even though there are significant differences in their precise realization in the two phrases. For instance, segment 1 is noticeably higher in frequency than segment 13 even though they appear as the first segment in both phrases and perform the same function (a high tone followed by a gradual descending contour). There are few moments of level pitch and many instances where there is an apparent peak (1), trough (7) or overshoot (9). Glides are also evident at (5), (17) and (24). Some of these glides are conditioned by linguistic determinants (see more detailed analysis of *Baba Wami* in Chapter 4 for discussion). What all these details substantiate, however, is the argument that the melody of this song is not reducible to a series of discrete pitches patterned in time, and that pitch targets and glides better reflect the gradience and variation of melody in general.

3.4 Pitch glides

Pitch glides are gradients that exist as autonomous categories independent of pitch targets. However, these glides need not be categorical, as we shall see in Chapter 4, and so establishing their function is important in the analysis. In linguistics a glide is usually referred to as 'contour tone,' but since the term 'contour' already has a clear meaning and literature in music studies I have preferred the more neutral term 'glide.' In music, gliding pitch is usually considered a non-essential embellishment of an existing pitch array. But in linguistics pitch glides may be phonological in tone languages where they provide a precise tonal signature for contrastive pitch categories (see discussion below). This point about glides as categories underscores the importance of an approach to song melody as prosody. Elements of both domains are integrated and virtually indistinguishable in sung utterances. Not so for instrumental music. This is because of the limitations of many instruments in producing non-discrete notes (especially keyboard and percussion), conventions of practice that emphasize discrete pitch categories (partly for ensemble purposes), and the absence of a semantic framework that requires unambiguous articulation of pitch contrasts for tone languages. Phonetic cues are mutually reinforcing when it comes to speech and song perception, and so glides must be integrated into a cognitive theory of melody. There is a close relation between pitch targets and glides since transitional glides to pitch targets are ubiquitous. Even where singers aim to reproduce 'discrete pitches'—as in Western art music—there is invariably a degree of gliding pitch between two articulatory gestures. These sorts of non-categorical glides do convey important perceptual information but it is important to show, as I do in the next chapter, that some glides operate as phonological categories independent of this phonetic detail. In sum, both glides and targets carry important pitch information for the listener, and *both* may operate as *combinatorial categories*.

3.5 Hierarchical structure

Hierarchical organization is a second element of musical melody thought to differentiate it from pitch patterning in speech. This is because speech melody is more or less continuous, unfolding without discrete pitch events or with few of them. It is structured differently from melody even if there are conventions of cadence or closure. Crucially, hierarchy has been studied as a series of pitch events that articulate a stable pitch structure. “For a series of stimuli to form separable events which can act as elements within a hierarchy, there must be some degree of closure. Closure—the arrival at relative stability—is a result of the action and interaction among the several parameters of music.”[152] Closure is important because it reinforces the sense of a tonic or stable pitch attractor within a series of pitches; it impresses a structure and hierarchy on a range of pitch events within a circumscribed durational frame. Many studies claim that hierarchical features are in part determined by the ordering of tonal features. “Music theory suggests that musical structure in tonal music can be accounted for by the hierarchies that apply to the musical elements and that each key is associated with a particular ordering of stability within the set of single tones and chords. A number of psychological investigations have obtained quantitative empirical support for the internalization of these hierarchies by listeners familiar with tonal music.”[153] Do we find such hierarchies in world music traditions or are these hierarchies specific to Western traditions? If we expect there to be strong homologies across musics then we must demonstrate a hierarchy of pitch relations, a hierarchy that can only function in this way if it has two main components: (1) a limited set of pitch categories or classes; and (2) a tonic or tonal center. Constructing a system of this sort makes a larger assumption: That the syntagmatic dimension is itself hierarchically structured or, in other words, that there is an implicit tonality operating across a linear pitch sequence.

[152] Meyer, *Explaining* Music, 81.
[153] Krumhansl and Kessler, “Tracing the Dynamic Changes,” 336.

The idea that common-sense categories and large-scale hierarchical forms are accurate representations of hierarchical organization is problematic, as was pointed out in chapters 1 and 2. Top-down approaches that impose global attributes tend to overlook perceptual constraints that are local. Meyer and Rosner point out conceptual problems with Schenkerian and generative accounts of hierarchical structure that rely on such global attributes:

> Analysts such as Schenker (1956) and Lerdahl and Jackendoff (1983) may well produce hierarchical diagrams of quite long musical passages. But if "higher" events exert increasingly tenuous perceptual influences, the psychological force of the outputs from such analytical machinery will dwindle quickly, whatever may be their value for the theory of music. In this regard, we must point out a fundamental difference between the tree structures used in linguistics and those presented by music theorists like Lerdahl and Jackendoff. The top node of a grammatical tree is an immediately observed datum: a sentence or an utterance. It represents some incident, occurring over time, which can be entered completely and rapidly into memory. The associated tree decomposes that uppermost node into parts at several lower levels of a strict hierarchy. The lowest nodes in music-theoretic tree structures, however, represent a datum: an actual stretch of music. Quite often, only fragments of it are held faithfully in memory. The lower nodes in the tree are not decompositions of higher ones. Instead, higher nodes are *selections* from among lower ones. We therefore cannot believe that the increasingly higher nodes, which represent ever more rarified selections, form the core of music perception.[154]

---

[154] Burton Rosner and Leonard B. Meyer, "The Perceptual Roles of Melodic Process, Contour, and Form." *Music Perception* 3, no. 1, (1986): 37.

Meyer and Rosner point to the gap between purported hierarchical structures and the "core of music perception." Like Narmour, they are skeptical of a totalizing theory that ignores the processual features of perception and cognition in real time. These observations are in keeping with the findings of empirical studies by Tillmann and Bigand (2004) and Cook (1987) who do not find support for the perception for large-scale tonal structures in Western music.[155]

Nicholas Cook is careful to show that global frameworks for understanding large-scale tonal closure are not grounded in cognitive reality, even if such theories do explain the organization of Classical and Romantic compositions. "Theorists as diverse as Schenker, Meyer, and Lerdahl and Jackendoff all explain the aesthetic effect of music in terms of hierarchical relationships between [ ] large-scale tonal structure and the local events of a given composition. Can it be empirically shown that the large-scale tonal closure of Classical and Romantic compositions plays a foundational role in listeners' aesthetic responses to the music?"[156]

> The direct influence of tonal closure on listeners' responses is relatively weak and is restricted to fairly short time spans – much shorter than the duration of most tonal compositions. Although large-scale tonal structure may not in itself be perceptible, it plays an important role as a means of

[155] It is important to recognize that hierarchies that are emergent are quite different to the sorts of analytical reductions we impose on our transcriptions of notated pitch events. It is likely that the mind-brain determines the relative weight of such hierarchies and events based on experience. That is, domain-general mechanisms of pattern-finding locate cues in the acoustic signal that are ordered according to culture-specific usage. This gives rise to the extraordinary diversity of pitch patterning, scales, and harmonies used by humans the world over.

[156] Nicholas Cook, "The perception of large-scale tonal closure," *Music Perception* 5, no. 2 (1987): 198.

> compositional organization, and it is argued that the theory of tonal music is more usefully regarded as a means of understanding such organization than as a means of making empirically verifiable predictions regarding the effects of music upon listeners.[157]

This perceptual salience of local as opposed to global structures has been demonstrated in experimental studies. Barbara Tillmann and Emmanuel Bigand report that, "[t]he musical brain, which appears to be very smart when dealing with fine musical structures in short temporal windows, turns out to be somewhat defective when short time structures have to be integrated in larger structures."[158] Still, even if listeners are focused on "short temporal windows," they do seem to continually construct expectations based on tonal contexts. More importantly, this bias is not entirely based on listener expertise or skill level alone.

> [O]verall the data suggest that local structures prevail over global structures for the listener – at least in musical pieces between twenty seconds and three minutes. Global structures seem to have only weak influences on perception, and local structures seem to be much more important. Independently of level of musical expertise, listeners have difficulties considering relations between events that are far apart in time. And yet, understanding such distant relations would be necessary to integrate events into an overall structural organization. When referring to the hierarchical conception of the sliding temporal window presented above, the data suggest that listeners do not connect, to any appreciable degree, the different windows at a higher hierarchical level. Consequently, perception seems to take place through a succession of local structures.[159]

---

[157] Ibid, 197.
[158] Barbara Tillmann and Emmanuel Bigand, "The Relative Importance of Local and Global Structures in Music Perception," Special Issue: Art, Mind, and Cognitive Science, *The Journal of Aesthetics and Art Criticism* 62, No. 2 (2004): 218.
[159] Ibid.

The finding that perception is based on a succession of local structures rather than global hierarchies seems to undermine generative and Schenkerian accounts that insist on a top-down model of music cognition. Narmour (1983) recognized several important problems with the concept of tree-structured hierarchies. He argued that, "levels exist in artifactual phenomena *only* as partially decomposable events. To conceive level-relationships in music as essentially a tree system of completely decomposable parts can only lead to unsatisfactory and erroneous analyses. Musical hierarchies are inherently networked both vertically and horizontally. It is just such complexity that makes the formation of music theory so difficult."[160] This idea that hierarchies are not decomposable is very important to the theory of prosodic structure developed here. It supports the idea of treating melody as a 'complex adaptive system' that is networked and which consists of pitch categories and relationships that are irreducible. The categories and structures of which they are a part must be understood as integrated, not decomposable. Narmour expands on these observations: "Now what distinguishes networked hierarchies from tree-structured ones is that in networks, connections between levels tend to be asymmetrical and nonsystematic."[161] The problem of establishing the relationships between levels in hierarchies that are partially independent and partially decomposable is of course complex and requires multidimensional scaling of parameters. This is at present beyond the scope of this study, but it remains an important goal for a theory of melody that takes into account the multiple components recognized in this chapter.

---

[160] Narmour, "Some Major Theoretical Problems Concerning the Concept of Hierarchy in the Analysis of Tonal Music," 196.

[161] Narmour, "Some Major Theoretical Problems Concerning the Concept of Hierarchy in the Analysis of Tonal Music," 149.

I have introduced several categories of pitch structure in this chapter, including pitch targets and glides. But there are also larger (syntagmatic) elements that are not reducible to such localized units (contour and intonation). If we understand these as elements in a network we see how their mutual interaction produces a complex system in which no single pitch element is given priority. For instance, many theories of tonality and hierarchical structure rely on a tonal center even though there are many other factors that may prevail. Glides may be used to establish closure, or intonation may provide a scaffold for the entire melodic sequence within a circumscribed tessitura. The case study that follows shows that establishing a single tonal center or tonic pitch and set of recursive, hierarchical structures based upon it, is not always possible. We may rethink melody as an emergent, usage-based phenomenon, rather than as a rigid set of tonal determinants. This also means exploring the set of possible combinatorial elements and how they are acquired. Studies of 'chunking' in language and music suggest that these elements are acquired based on their functional salience (or repetition) within a practice.

3.6 Chunking

According to Newell (1990), "A chunk is a unit of memory organization, formed by bringing together a set of already formed chunks in memory and welding them together into a larger unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organization of memory. Chunking appears to be a ubiquitous feature of human memory."[162] Crucial to the process of chunking is repetition.

[162] Quoted in Bybee, *Language Usage and Cognition*, 34.

Humans tend to learn various sizes of chunks as a result of repetitions experienced in everyday speech. These chunks may include a single word, an idiom, a phrase or even a sentence (observations that accord with the action of Tomasello's domain-general learning mechanisms for pattern recognition discussed in the previous chapter). Newell's statement suggests that chunking is dependent on prior knowledge of the constituent elements of chunking. In other words, to chunk a phrase requires that one first has access to the individual words, or to chunk a word implies knowledge of phonemes. But studies have shown that the process may proceed in reverse from whole chunks to parts thereof.

> While language users constantly acquire more and larger chunks of language, it is not the case that in general the language acquisition process proceeds by moving from the lowest level chunks to the highest. Even if children start with single words, words themselves are composed of smaller chunks (either morphemes or phonetic sequences), which only later may be analyzed by the young language user. In addition, however, children can acquire larger multi-word chunks without knowing their internal composition (Peters 1983). The acquisition process in such cases consists of the analysis of such a chunk into smaller units and a growing understanding of what parts of the chunk are substitutable or modifiable.[163]

Combinations of chunks vary in size but are not dependent on smaller, discrete units as the building blocks for larger chunks. This suggests that humans fixate on repeated, recognizable, and *holistic* elements that have communicative salience and memorize them. These may be inserted into combinations structured by generic language-specific

---

[163] Bybee, *Language Usage and* Cognition, 35).

constraints. I propose that chunking of this sort in language is analogous to the chunking of melodies in song. There is preliminary evidence that similar cases exist for chunking in sight-reading and the memorization of melodies.[164] Chunking may apply to pitch targets, glides or even contours. If we take music to be a flexible, complex system, then we see how combinatorial components may well be adapted or reconfigured to include chunks of this sort. This reinforces the idea that hierarchies are networks rather than exclusively top-down or bottom-up processes. Domain-general mechanisms of pattern-finding provide the means for processing diverse pitch phenomena, including targets, glides and other 'chunks' of material. Given the attentional and memory constraints on the perception of melody in real-time we need to return to processual studies.

3.7 Contour

Most music theorists take 'melody' to be a sequence of discrete pitches patterned in time.[165] Contour is generally treated as a conglomerate of these localized pitch events rather than as a holistic, singular utterance. Both must factor into a cognitive theory of song melody where gradient pitch phenomena are not reducible to individual discrete pitches. How do non-adjacent notes relate to one another and how does contour shape design? How can discrete combinatorial models account for the constraints imposed by continuously varying pitch contours in song? Since I have already discussed pitch

---

[164] But several studies on the role of chunking in memorization and musical form are valuable. See: Christiane Neuhaus, "Processing musical form: Behavioral and neurocognitive approaches. *Musicae Scientiae* 17, no. 1 (2013): 109-127; Susan Wheatley, *An application of chunking to the memory and performance of melodic patterns* (PhD dissertation: University of Michigan, 1991).
[165] Mark A. Schmuckler, "Melodic Contour Similarity Using Folk Melodies," *Music Perception: An Interdisciplinary Journal* 28, no. 2, (2010): 169-194.

categories at length in this chapter I will review studies of music contour only briefly. This is because most studies of contour emphasize discrete pitch elements anyway. Following this section I develop a model that integrates my concepts of pitch targets and glides with an important linguistic feature: intonation.

Early cognitive research in music emphasized the role of holistic pitch contours in cognition. Erich M. von Hornbostel, influenced as he was by Gestalt psychology, drew attention to such features in his 1928 essay on African music.

> The singer does not possess in his mind a store of notes or intervals from which he selects and combines those which may please his ear. What he really aims at more or less consciously is melody forming an undivided unity which he performs at one stroke as an athlete does an exercise. This melodic unity determines the parts which we afterwards dissect in the course of our analysis. Recent experimental investigations have shown how, and in what way, the size of intervals in unaccompanied song depends, even with our singers on the 'form' (*Gestalt*) of a melody.[166]

Von Hornbostel argues that the gestural or holistic qualities of melody are crucial to understanding perception and cognition. The relationship to intonation contours that is more important than the "store of notes or intervals" that we so often assume to be categorical. These holistic features are not discrete and seem to be chunked much in the ways discussed in 3.6. This approach to contour contrasts markedly with some recent

---

[166] Erich M. von Hornbostel, "African Negro Music," *Africa: Journal of the International African Institute* 1, No. 1 (1928): 35.

work in music cognition that has sought out the relationships between discrete pitch events rather than the gradience of combined categories.

Scholars in music cognition have developed experimental models to account for melodic contour.[167] Edworthy and Quinn, in two separate studies, both found that contour relations supersede the role of adjacent pitch relations in the perception of tonal melodies.[168] Quinn's experiments showed that "each note of the melody is judged relative to its pitch relations with all other notes in the melody."[169] These findings suggest that instrumental melodies are organized according to a system of tonality that is hierarchically structured. Missing from this account are the gradient elements so characteristic of the contours of song melodies. Furthermore, these findings tacitly assume a synoptic rather than a processual model of cognition (see Cone 1968; Chapter 2).[170] Evidence for the primacy of contour as a holistic phenomenon is provided by Edworthy (1985), who found that contour is processed *before* individual notes are encoded, a finding that replicates work on infant-directed speech (see Chapter 1). This takes us back to Hornbostel's emphasis on melody as a Gestalt-like phenomenon. Certainly there are alternative conceptions for this in the work of 'schema' studies by Gjerdingen, Dowling, Harwood, and others, though here too the emphasis is on the

---

[167] For experimental models see: Elizabeth West Marvin and Paul A. Laprade, "Relating Musical Contours: Extensions of a Theory for Contour," *Journal of Music Theory* 31, no. 2 (1987): 225-267; Ian Quinn, "The combinatorial model of pitch contour," *Music Perception* 16, no. 4 (1999): 439-456; Mark A. Schmuckler, "Components of melodic processing," in *Oxford Handbook of Music Psychology*, ed. Susan Hallam, Ian Cross, and Michael Thaut (Oxford: Oxford University Press, 2009): 93-106.
[168] Edworthy (1985); Quinn (1999).
[169] Schmuckler (2010, 170).
[170] Edward T. Cone. 1968. *Musical form and musical performance*. New York: Norton.

discrete pitch arrays of notated (and instrumental) music rather than the sorts of gradual and continuous pitch contours characteristic of human vocalizations.[171] It is therefore primarily to prosody studies that I turn in order to model contour and other gradient-like features in song melodies. Fortunately many linguistic models of pitch are not predicated on discrete intervals or categories. The challenge, therefore, is to integrate the very broad analytical categories for functional units in linguistics with the precise categories of music theory.

PART 2: PROSODY

3.8 Phonetics and Phonology

In linguistics, the study of prosody is the domain of both phonetics and phonology, two modes of inquiry that are increasingly integrated in theory and practice. For the remainder of this chapter and the next I refer to literature on both.

"Phonetics deals with the production (*articulatory phonetics*), transmission (*acoustic phonetics*), and perception (sometimes *auditory phonetics*) of speech sounds. Somewhat more narrowly, it is often seen as being centrally concerned with delimiting the features that can serve as the basis of linguistic distinctions."[172] These *physical* properties are studied using precise and usually quantitative instrumental measures. Still, it is important

---

[171] W. Jay Dowling and Dale L. Harwood, *Music Cognition* (Orlando: Academic Press, 1986); Gjerdingen, *A Classic Turn of Phrase;* Robert Gjerdingen, "Categorization of Musical Patterns by Self-Organizing Neuron-like Networks," *Music Perception* 7, no. 4 (1990): 339-370.

[172] D. Robert Ladd, "An Integrated View of Phonetics, Phonology, and Prosody," in *Language, Music, and the Brain: A Mysterious Relationship*, ed. Michael Arbib (Cambridge, MA: MIT Press, 2013): 274.

to recognize that phonetics has a social dimension. John Laver defines it as "any learnable aspect of use of the vocal apparatus" and points out that "the phonetic level of description is abstract, not concrete" because it emphasizes the relative sameness of sounds across a speaker population in which there are organic differences between individuals.[173] A phonetic description "does not require knowledge about the formal, linguistic value that the event being described might have as a coded, communicative element in some particular language. In this sense, phonetic description of a given stretch of speech is held to be independent of the phonological description of the language involved. This forms the basis on which descriptive phonetic theory can be regarded as a general theory capable of application to the sounds of any language in the world."[174]

Phonetics is sometimes considered to be outside the scope of mainstream linguistic research, especially in the American structuralist and generative traditions, because it focuses on empirical studies that quantify physical properties of speech phenomena rather than their function. In contrast, "phonology deals with the way sound is structured in language. The most basic aspect of this structure is the *phonemic principle*. Since the late nineteenth century it has been clearly understood that one language may signal semantic distinctions by means of phonetic differences that are irrelevant in another language […]. This means that knowing a language involves knowing both a system of phonetic

---

[173] John Laver, *Principles of phonetics* (Cambridge University Press, 1994): 28. "To say that two sounds are phonetically equivalent rests on an idealizing assumption that organic differences between speakers can be ignored in evaluating phonetic quality, as if both speakers could be held to be producing their performance on the same notional vocal apparatus" (Ibid, 29).
[174] Ibid.

categories and the abstract structural relations between the categories."[175] Phoneticians study the physical properties of sounds and their perception whereas phonologists explain the organization of abstract sound categories and their structural regularities. "[T]he function of phonology is to relate the phonetic events of speech to grammatical units operating at the morphological, lexical, syntactic and semantic levels of language."[176] This study uses methods of phonetics to describe and analyze the properties of pitch prosody (Chapter 4). But the theoretical position developed here is phonological in the sense that it offers an explanation for how these units are organized and networked into a complex adaptive system. Phonetics and phonology are complimentary approaches to the study of speech sounds and structure. Categories of speech prosody are not self-evident and their properties require careful phonetic study.

3.9 Prosodic features and prosodic structure

This study is titled 'melody as prosody' and so it is very important that we understand what 'prosody' entails. In linguistics the term *prosody* has been used for many different features, including length, accent, stress, tone and intonation.[177] Delimiting this set is difficult because of the wide range of prosodic features and the fact that none operate in isolation. In general terms we observe that prosody refers to the *suprasegmental* features of language. The consonants and vowels that make up words are its segmental features.

---

[175] Ibid, 275.
[176] Ibid, 30.
[177] The word prosody derives from the Greek 'prosodía' "which appears to signify something like 'song sung to music' or 'sung accompaniment', implying that prosody is the musical accompaniment to the words themselves." Anthony J. Fox, *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals* (New York: Oxford, 2000): 1.

Patterns of relationships that are relative and contrastive over two segments, a phrase or the utterance as a whole must be classed separately. The relationship of these prosodic features to segments is complex and is the subject of much debate. Structuralists regard suprasegmentals as superimposed 'on top of' segments. In classic generative phonology, on the other hand, they are attributes *of* segmental units. More recent non-linear approaches treat suprasegmentals as complex, multidimensional systems, an approach that I will draw on in this dissertation.[178] In autosegmental metrical theory prosodic features occupy their own autonomous tier that is entirely independent of segmental features. What all of these approaches agree on is that there is an essential syntagmatic dimension to prosodic features and that contrastive relations exist beyond the segment.

There are both phonetic and phonological criteria for describing prosodic features. From a phonetic perspective we observe three main dimensions of activity in the vocal tract: laryngeal, subglottal and supralaryngeal.[179] Most prosodic features are produced by laryngeal or subglottal activity whereas all segmental features are supralaryngeal. Note that pitch production, including the features of tone and intonation, is entirely controlled by the laryngeal muscles.[180] This physiological dimension suggests that, "we could consider prosodic features to be […] more fundamental, in the sense that segmental features involve the modification of an air-stream which is already specified for prosodic

[178] Ibid, 2.
[179] For further discussion of these mechanisms see Laver, *Principles of Phonetics*.
[180] Articulators segment the airstream produced by the laryngeal muscles and alter its expression in important ways.

features."[181] The phonological basis for suprasegmentals is different. The most important aspect is that these features apply to domains larger than the individual segment and that their linguistic dimensions are both paradigmatic *and* syntagmatic.[182]

Palmer and Hutchins (2006) define prosody as "acoustic changes in frequency, amplitude, and duration that form grouping, prominence, and intonation."[183] Prosodic features assist listeners in disambiguating the meaning of words and phrases and in interpreting the illocutionary intent (bearing or disposition) and affective state of speakers. In other words, prosodic features perform both linguistic *and* paralinguistic operations and distinguishing one from the other can be difficult. All these attributes are found in song prosody. The only real difference with speech is that pitch phenomena in song are arguably more complex in structure. *My study does not therefore distinguish 'musical prosody' from linguistic prosody, because the evidence I have considered suggests that song has the same prosodic features as speech*.

It is important to recognize that studies of musical and linguistic prosody, though related, employ similar terminology for subtly different phenomena. Prosodic terms are not interchangeable in music theory and linguistics. They serve different *theoretical* agendas in their respective literatures, even if there are distinct resemblances between music and

[181] Ibid, 4. This observation is potentially very interesting for studies of origins and evolution.
[182] 'Suprasegmental' is not equivalent to 'prosodic' because the latter may be treated both segmentally and nonsegmentally.
[183] Caroline Palmer and Sean Hutchins, "What is musical prosody?" In *The Psychology of Learning and Motivation* 46, ed. B.H. Ross (2006): 246.

language in the *physical* properties of melody, tone, and intonation in the auditory signal. Several terms are commonly applied to the domains of song and speech that have very different meanings in music studies and linguistics. *Melody*, for instance, is used to describe the pitch patterning of speech, song, and instrumental music, depending on context, and is often equated with intonation in speech. Frick states that, “A prosodic contour is analogous to a simple melody with dynamics in music”.[184] Consider also the following monographs on speech prosody: *The Melody of Language: Intonation and Prosody*,[185] *A perceptual study of intonation: An experimental-phonetic approach to speech melody*,[186] and *The Music of Everyday Speech: Prosody and Discourse Analysis*.[187] Sieb Nooteboom’s review of speech prosody literature—aptly titled, ‘The prosody of speech: melody and rhythm’—reminds us that the word ‘prosody’ itself has a musical etymology derived from Greek, “where it was used for a song sung with instrumental music.”[188] *Rhythm and meter*, also key concepts in both domains, are used interchangeably in linguistics but not in music. Some linguists have been influenced by such cross-disciplinary borrowing of terms and concepts and have used notions of meter

[184] Robert W. Frick, “Communicating Emotion: The Role of Prosodic Features,” *Psychological Bulletin* 97 no. 3 (1985): 421.
[185] Linda R. Waugh and Cornelis H. van Schooneveld eds., *The Melody of Language: Prosodics and Intonation* (Baltimore, MD: University Park Press, 1980).
[186] Johan ‘t Hart, René Collier, and Antonie Cohen, *A Perceptual Study of Intonation: An Experimental-Phonetic Approach To Speech Melody* (Cambridge: Cambridge University Press, 1990).
[187] Ann Wennerstrom, *The Music of Everyday Speech: Prosody and Discourse Analysis* (Oxford: Oxford University Press, 2001).
[188] Sieb Nooteboom, “The Prosody of Speech: Melody and Rhythm,” in *Handbook of the Phonetic Sciences* ed. W. Hardcastle and J. Laver (Oxford: Blackwell, 1997): 640.

and melody in work on language.[189] This has been an important feature of studies of motherese, or infant-directed speech.

The three principle prosodic parameters are pitch, amplitude, and duration. These are suprasegmental features that are shared by music and language. Many scholars have drawn on elements of one to explain the other. The first is music acquisition: ‘motherese,’ or infant-directed speech (IDS), is taken to be a cross-cultural universal that seems to consist of features intrinsic to both song and speech. As we observed in Chapter 1, various evolutionary scenarios have been proposed to account for the ‘universal’ qualities of prosody.

A second area of study is emotion in music. How is affect conveyed through the expressive features of musical performance? Subtle (gradient) changes in timing, accent, and intonation show how musicians use prosodic features for expressive purposes. A third key area is the study of musical performance itself. Expressive timing, melodic contour, and various forms of rhythmic perturbation have been considered important in directing listeners attention to important *structural* features of a particular work. These various approaches to musical prosody are increasingly interrelated and cross-disciplinary, including research in music theory, linguistics, cognitive neuroscience and psychology.

---

[189] Mark Liberman and Alan Prince, “On Stress and Linguistic Rhythm,” *Linguistic Inquiry* 8, no. 2 (1977): 249-336.

What I draw attention to in this chapter are the important conceptual linkages between the organization and structure of pitch patterning in speech and song. This requires a careful articulation of the concepts already identified in the first part of the chapter with two main elements of speech prosody: tone and intonation. One of the main challenges for studies of musical prosody is to establish a clear conceptual apparatus for investigating the purposes of prosodic features across domains while factoring in these similarities.

3.10 Pitch phenomena in language

There are three main linguistic determinants of pitch structure: accent, tone and intonation. One useful way of distinguishing the domain of pitch accent from tone is through recourse to Ferdinand de Saussure's terms *syntagmatic* and *associative* (or *paradigmatic* in Roman Jakobson's later, and now conventional, formulation).[190] In his *Course in General Linguistics* Saussure designated the consecutive linear combination of two or more elements as syntagms. These stand in opposition to one another and all other linear elements in an utterance. Associative or paradigmatic elements, on the other hand, consist of the set of simultaneous potentials for any single unit in an utterance. "[The] syntagmatic relation is *in praesentia.* It is based on two or more terms that occur in an effective series. Against this, the associative relation unites terms *in absentia* in a

---

[190] These distinctions have been taken up in music semiotics by several scholars. See: Nicolas Ruwet and Mark Everist, "Methods of Analysis in Musicology," *Music Analysis* 6, no. 1/2, (1987): 3-9+11-36; Philip Tagg, "Analysing Popular Music: Theory, Method and Practice," *Popular Music* 2 (1982): 37-67; Patrick McCreless, "Syntagmatics and Paradigmatics: Some Implications for the Analysis of Chromaticism in Tonal Music," *Music Theory Spectrum* 13, No. 2 (1991): 147-178.

potential mnemonic series."[191] "Whereas a syntagm immediately suggests an order of succession and a fixed number of elements, terms in an associative family occur neither in fixed numbers nor in a definite order. If we associate *painful, delightful, frightful*, etc. we are unable to predict the number of words that the memory will suggest or the order in which they will appear. A particular word is like the center of a constellation; it is the point of convergence of an indefinite number of co-ordinated terms."[192] Structuralist linguists have tended to focus on the potential substitution of paradigmatic elements rather than their associative functions (recalling ideas stored in memory that are not necessarily interchangeable).

Pitch accent gives prominence to one or another element in a linguistic sequence or utterance. It has a contrastive function and is therefore a syntagmatic feature. Tone, on the other hand, has "lexical or grammatical significance as an intrinsic property of a morpheme, word, or grammatical construction."[193] It is a paradigmatic feature. In other words, pitch accent is not an inherent property of a particular phoneme or word but is used to stress that element in relation to others in a linear sequence. A speech tone cannot be varied in this way because its relative pitch height is intrinsic to its meaning and function. These two elements of prosodic structure operate at the level of individual segments even if they do so relationally rather than autonomously (hence their suprasegmental designation). Intonation is different, since it has a 'global' function and

[191] Ferdinand de Saussure, *Course in General Linguistics*, transl. by Wade Baskin, ed. by Perry Meisel and Haun Saussy (New York: Columbia University Press, 2011): 123.
[192] Ibid, 126.
[193] Fox, *Prosodic Features and Prosodic Structure*, 179.

has an entirely grade or continuous pitch realization. Intonation is intrinsic to all voiced utterances while pitch accent and tone vary from one language to the next. Intonation encompasses an entire phrase or sentence and its function is "discourse-oriented rather than lexical or grammatical."[194] I focus on speech tone and intonation in the remainder of this chapter since these are the categories most relevant to my analyses in Chapter 4.

Given that pitch features—and intonation in particular—are among the most contested theoretical domains in linguistics, this means choosing between a wide range of largely incompatible positions. On a phonetic and descriptive level there are several methods for analysis; but there is little agreement in studies of classic generative phonology, because both tone and intonation are difficult to segment given the continuously varying, linear nature of pitch patterning in all speech. In the sections that follow we will consider linguistic conceptions of 'tone' and 'intonation' in turn.

3.11 Tone

Modern studies of tone begin in effect with Kenneth Pike's *Tone Languages* (1948).[195] Pike defined a tone language as "a language having lexically significant, contrastive, but relative pitch on each syllable."[196] First, consider the phonetic basis for tone: "Lexical pitch variations and intonational pitch variations are phonologically represented as tones, like H(igh) and L(ow), which form a string of elements running parallel to the string of

---

[194] Ibid.

[195] Kenneth L. Pike, *Tone Languages: A technique for determining the number and type of pitch contrasts in a language, with studies in tonemic substitution and fusion* (Ann Arbor: University of Michigan Press, 1948).

[196] Pike, *Tone* Languages, 3.

vowels and consonants. Like vowels and consonants, tones may delete, assimilate, or change their value in particular contexts."[197] Unlike notes in a melody, tones are contrastive and have relative rather than precise and consistent pitch targets. Thus many linguists differentiate between contrasting high and low tones. But the analysis of tone is not straightforward. In many tone languages there may be intermediate levels and contours, and tones may also be masked by concatenation of syllables, by pitch accents and by intonation contours.

Contours (glides) are a common feature of some tone systems where speakers glide from one pitch target to the next. What is significant about these glides is that they are not simply ornamental but are meaningful units in and of themselves. "[T]he distinction between level and gliding tones has been regarded as a fundamental dichotomy between tones, and hence between tone-systems. This implies that contour tones are a legitimate type of tone, different from, but of equal status with, level tones, and many scholars have made this assumption."[198] There is considerable debate about the nature of contour tones, largely as a result of different theoretical frameworks. Still, it cannot be denied that high and low contrastive tones are insufficient to describe the complexity of tone categories used in some languages.

> If a language places a heavy information load on its tonal contrasts, level tones do not suffice. There are many languages that contrast four or more tones, and, as we have seen, this number of

---

[197] Carlos Gussenhoven, *The Phonology of Tone and Intonation* (New York: Oxford, 2004): xvii.
[198] Fox, *Prosodic Features and Prosodic Structure*, 194.

> contrasting levels is rare. Instead, a smaller number of level tones is supplemented by the addition of one or more contours. These may be simple rises or falls, or tones which first rise and then fall (convex tones) or first fall and then rise (concave, or dipping, tones). The existence of the more complex tones usually implies the existence of simple rises and falls. Falls are much more common than rises (Zhang 2000, 2001).[199]

Contour tones often have pitch targets. They do not, therefore, completely nullify a model of melody that holds to combinatorial principles and categories. We see this in Zulu song prosody, discussed at length in the case study in Chapter 4, where a continuously varying pitch contour has consistent peaks, troughs, and occasional level pitch. It is an interactive system that accommodates the features of speech tone within a combinatorial framework.

This brings us to another feature of most tone systems: *downstep*. This is the phenomenon in which a high tone is followed by one or more successive high tones with minimal descending contrast in pitch height. In other words, a series of syllables with consecutive high tones results in a slight lowering of pitch on each syllable rather than a series of larger leaps. This is not to be confused with *downdrift*, a feature of intonation that is common but not universal and involves a gradual lowering of pitch over the course of the entire phrase or sentence. In tone languages downstep is constrained by the requirements of tonal contrast (high vs. low). In most instances the realization of the interval between a high and contrastive low tone is reduced by this phenomenon.

---

[199] Moira Yip, *Tone* (Cambridge: Cambridge University Press, 2002): 27.

Basic descriptors of this sort are sufficient for a phonetic analysis, but phonological properties are more complex, because they must be assigned within the context of a particular tone-system. Several factors are relevant, including: domain, number, and kind of tones. The 'domain' is the unit of speech. Pike and most others regard this as the syllable, though the issue is trickier for languages like isiZulu, because speech tone categories appear to operate across and within syllables. 'Number' and 'kind,' are simpler: for instance, most Bantu languages have two contrastive level tones, High and Low, as well as glides. We may also distinguish between register and contour systems. IsiZulu is unlike many other Bantu languages in that it possesses both level tones and glides. I will not attempt a complete theory of prosodic structure, because each constituent element is a complex topic that would require separate treatment. What I point to here is the fact that each of these must be accounted for in a cognitive theory of melody, and that the gradient features of prosody undermine the discrete combinatorial models in common use. This is especially true of intonation.

The idea that tones should be understood as binary or unary features is a foundational tenet of the autosegmental-metrical theory that I will consider in more detail below. This idea stems from distinctive feature theory developed by Jakobson, Fant and Halle (1952) and later Chomsky and Halle (1968).[200] Moira Yip describes the contemporary

[200] Roman Jakobson, C. Gunnar Fant and Morris Halle, *Preliminaries to Speech Analysis*. *The distinctive features and their correlates* (Cambridge, MA: MIT Press, 1952); Noam Chomsky and Morris Halle, *The Sound Pattern of English* (1968, Repr. Cambridge, MA: MIT Press, 1991).

understanding of phonological treatments of tone in terms of distinctive features as follows:

> Phonological representations are categorical, using either binary or unary features. It is the business of the phonology to generate an output out of these elements, in which most segments are specified for most features, but some may lack specifications for certain features. In particular, some syllables may lack tones at the end of the phonology. The phonetics then interprets this phonological output, making use of all phonological information: featural, structural, phrasal, and so on. This phonetic component ultimately produces instructions to the articulators; these instructions may or may not be binary, but in any case they result in a continuous signal in which every syllable is pronounced at some fundamental frequency. In this final acoustic output, the pitch values of the successive syllables will not be limited to a few binary distinctions, but will cover the range of the speaker's voice. The phonetic output is thus often termed 'gradient.'[201]

The fact that phonological representations use either binary or unary features contrasts with the 'gradient' phonetic output. This is the crucial difference that I wish to highlight in this chapter and the next. Phonological distinctive features tend to be classified as binary or discrete digital outputs, whereas categories used in song tend toward gradience. Autosegmental metrical studies of intonational phonology have also focused on binary contrasts, in this way extending a conceptual framework developed in generative linguistics. But there are other possibilities for studying intonational and other gradient phenomena that have developed by scholars of perception.

---

[201] Yip, *Tone*, 10-11.

3.12 Intonation

Intonation is a feature of all languages and is a basic element of prosodic structure. It is characterized by a continuously unfolding pitch contour that, despite momentary breaks for unvoiced consonants, is perceived as an unbroken sequence. These 'global' trends are post-lexical and contrast with the localized lexical properties of speech tone. Intonation is meaningful in and of itself. An intonation contour may signal a question, an agreement, or a sense of confusion. It also plays a vital role in communicating paralinguistic meanings, including affect. To incorporate intonation into a theory of melody requires that we establish its meaning within a specific communicative context. Like other prosodic features it does not operate as a consistent constraint or cognitive factor in every utterance; there is no possibility for a totalizing theory of intonational structure. This is why any theory of melody must treat song as a complex adaptive system that is continuously melding itself to context-specific requirements.

Anthony Fox writes that intonation has "a number of characteristics which set it apart from other prosodic features. In the first place, it is *meaningful*. Other features, both prosodic and non-prosodic, do not in themselves *have* meaning, but merely serve to distinguish meaningfully different linguistic items."[202] *Paralinguistic* meanings accrue from the affective valence associated with linguistic gestures that are not always under conscious control. In some cases these are akin to the affective states invoked in musical

---

[202] Fox, *Prosodic Features and Prosodic* Structure, 270.

pitch contours.[203] Paralinguistic features have been recognized to be a fundamental constituent of the speech signal since the earliest studies of prosody. “The central difference between paralinguistic and linguistic messages resides in the *quantal* or *categorical* structure of linguistic signaling, and the *scalar* or gradient nature of paralanguage.”[204] These gradient properties of paralanguage are intrinsic to the signal and operate within the bounds of phonological categories rather than across them. To signal emotion requires microprosodic shifts rather than large-scale fluctuations that would alter semantic meanings. The concurrence of the linguistic and paralinguistic dimensions of intonation reinforces the point that pitch patterning in song operates as multi-dimensional complex, communicating a dense informational load through a single parameter. The structure of intonation has been studied from several vantage points and in order to elaborate its role we must investigate its phonological structure.

[203] For discussion of how prosodic features of speech communicate emotions see: Frick (1985). For a recent study of contour and emotion in music see: Emery Schubert, “Modeling Perceived Emotion With Continuous Musical Features,” *Music Perception* 21, no. 4 (2004): 561-585. This study, like so many others (e.g. Edworthy 1985 and Quinn 1999 discussed above), uses discrete pitch elements rather than pitch gradients or ‘sentence-level’ features to model intonation. “[T]he aim of the present study was to examine how well emotional response could be mathematically modeled by using only a handful of interval-scale musical features” (580). The emphasis on intervallic structures is justified for the study Western instrumental art music, but the findings of such studies do not generalize to studies of song.

[204] D. Robert Ladd, *Intonational Phonology*, Second Edition (Cambridge: Cambridge University Press, 2008): 37.

3.13 Intonational structure

Intonation is characterized by a continuously unfolding pitch contour rather than a series of discrete pitch events. This contour may be decomposed into a series of events or elements that constitute its phonological structure.[205]

> There is no doubt […] that the most important phonetic parameter involved in intonation is pitch. Phonetically, therefore, we can regard the intonation of an utterance as a continuous – and continuously varying – pitch pattern. There are, strictly speaking, breaks in this pattern for voiceless consonants, since there can be no pitch without voice, but these are not generally perceived as such; to the ear the pattern is uninterrupted (Laver 1994, 484). Since the pitch is continually changing – sustained level pitch is not common – the issue in the phonological analysis of intonation is usually a matter of identifying phonologically significant falls and rises in pitch, and determining the structure of the patterns produced.[206]

'Sustained level pitch' is rare in speech but it is also rare in many song cultures (see Chapter 4). If we are to extract pitch events from this 'continuously varying' pitch pattern then we need to focus on 'pitch targets' (including 'levels') and 'glides.' There are

---

[205] There are several parameters that may be invoked to convey prominence in an utterance. These include pitch, duration, and stress, and in most instances these three align. A single pitch nucleus may be used to give prominence to a single word or syllable in an utterance thereby changing the meaning of the sentence. Consider the following:

Mandela died today	Mandela died today
_ ^ - - -	_ _ _ _--^

The first statement informs us of Mandela's death whereas the second expresses a sense of incredulity: did Mandela really die today? These fluctuations in prominence may be lexical or sentence level and are also significant in song. The important point for the argument developed here is that prominence acts independently of melodic structure. Fluctuations of this sort are *not* categorical.

[206] Fox, *Prosodic Features and Prosodic Structure*, 274.

several models of intonational structure in linguistics that serve as a basis for investigations of this sort.

Early studies described intonation contours as whole 'tunes' and classified these by overall shape. But a simplistic classificatory system was later abandoned when scholars recognized that contours are *structured* and should not be studied as indivisible wholes. There are local elements that contribute to global structure, as Anthony Fox (2000) points out: "In the first place, there are many different patterns, with both similarities and differences between them, and these differences and similarities are localizable in specific parts of the pattern. Standard principles of linguistic analysis demand that we establish a structure with substitutable parts. Second, there must also at least be points in the pattern which must be aligned with appropriate parts of the utterance, and with other prosodic features. All of this suggests that the intonation pattern has a structure, and that it can be broken down into parts which have some degree of independence."[207] The major theoretical disagreements have to do with how best to characterize the constitutive elements of this structure. Should these be understood as discrete pitch events or as gradients? How localized should such events be: at the level of the phoneme, the sub-phrase, or the utterance? Competing theories propose an array of basic elements that range from 'pitch phonemes' and 'levels' to 'nuclear tones,' 'movements,'

[207] Fox, *Prosodic Features and Prosodic Structure*, 292.

'configurations' and 'accents.' I will not review this literature here.[208] Instead I focus on the two positions that are most relevant to our consideration of song prosody: the analysis of intonation in terms of 'pitch accents' (or levels) versus 'pitch movements' (or contours).

3.14 Levels

In his influential book *Intonational Phonology* (2nd edition, 2008), D. Robert Ladd outlines what he regards as a consensus position in recent studies of intonation. This is reflected in his definition of the term itself: "[Intonation is] the use of *suprasegmental* phonetic features to convey 'postlexical' or *sentence-level* pragmatic meanings in a *linguistically structured* way."[209] 'Sentence-level' or 'postlexical' refers to the fact that "intonation conveys meanings that apply to phrases or utterances as a whole" that exclude "features of stress, accent, and tone that are determined in the lexicon."[210] Finally, 'linguistically structured' describes the fact that "intonational features are organized in terms of categorically distinct entities (e.g. low tone or boundary rise) and relations (e.g. stronger than/weaker than). They exclude 'paralinguistic' features, in which continuously variable physical parameters (e.g. tempo and loudness) directly signal continuously variable states of the speaker (e.g. degree of involvement or

---

[208] See Janet Pierrehumbert, *The Phonology and Phonetics of English Intonation* (PhD: MIT, 1980). Fox, *Prosodic Features and Prosodic Structure*, Chapter 5, provides a detailed review and comparison of these different approaches.
[209] Ladd, *Intonational Phonology*, 4.
[210] Ladd, *Intonational Phonology*, 6.

arousal)."[211] Ladd's approach to intonation is characterized by an emphasis on intonation as a *sequential* tonal structure that comprises a series of localized *pitch accents* and *edge tones*. Global features are the product of local events.

This non-linear framework was first developed by Janet Pierrehumbert (1980) and has been widely adopted in the United States in particular, especially in the study of English intonation for which the Tobi (tones and break indices) system has been developed.[212] With regard to structure, "the intonation pattern is analyzed in terms of *pitch*-accents; each pattern thus consists of one or more such accents, which may in turn consist of either one or two tones. The tones are restricted to a binary H vs. L."[213] This system of classifying tones builds on the autosegmental metrical (AM) framework developed for African tone systems by Goldsmith, and used subsequently for the treatment of tone in many languages.[214] "AM theory adopts the phonological goal of being able to characterize contours adequately in terms of a string of categorically distinct elements,

---

[211] Ladd, *Intonational Phonology*, 6.

[212] See Ladd (2008) for an introduction to this system.

[213] Fox, *Intonational Phonology*, 296.

[214] See: John Goldsmith, *Autosegmental and Metrical Phonology* (Oxford: Blackwell, 1990). Goldsmith's PhD dissertation (1976) at MIT is the basis for this reference text, and much later work in the field. His theory emerged from the classic generative phonology then practiced at MIT. Given the critique leveled in the previous chapter against generative theory in linguistics and music theory, it is worth pointing out that the AM approach is "a direct continuation of the traditional work of generative grammar that was codified in Chomsky and Halle's *Sound Pattern of English* (SPE) in 1968, and which defined to a large extent the nature of the questions that were central to phonological theory in the ten- to fifteen-year period beginning in 1965 […] and extending into the late 1970s, the era of what we may be permitted to call 'classical generative phonology'" (Goldsmith 1990, 1-2). This program is of course opposed to the usage-based framework developed in this dissertation. And there is a further point to this critique: the tendency to reduce complex pitch phenomena to binary or unary bits in intonational phonology (Ladd 2008). This approach denies gradient phenomena, such as pitch contours, and is based on a mechanistic understanding of the relationship of phonetics to phonology. It also presupposes a computational mind for which there is no apparent evidence in studies of music cognition at present. To make such an assumption is contrary to the empirical evidence gathered in many studies (see Tomasello 2003 for a review).

and the phonetic goal of providing a mapping from phonological elements to continuous acoustic parameters."[215] It is the most widely practiced approach to intonation today. But there are reasons to be critical of its most basic premise: that intonational phonology should reduce pitch contours to a series of pitch accents or discrete events. Should intonation be understood primarily in terms of discrete contrastive pitch accents and tones, or is there room for gradient features too?

There are two main problems with AM theory that I will discuss here. The first is that it reduces all pitch patterning to binary contrastive categories that eliminate gradient features. This is untenable given the complexity of paradigmatic elements of melody. The problem is of course less important in non-tone languages like English in which there are no lexical pitch distinctions. A second problem with AM theory is that it lacks descriptive accuracy. According to the AM theory, "tonal structure consists of a string of local *events* associated with certain points in the segmental string. Between such events the pitch contour is phonologically unspecified and can be described in terms of *transitions* from one event to the next."[216] Notice how this focus on events emphasizes local *segments* at the expense of gradient features and what are explained away as 'transitions.' Pitch features that are not contrastive are considered superfluous to the theory. This may be a defensible position for languages like English, but it is *not* for most tone languages. In Chapter 1 I pointed out that song has a far more complex paradigmatic pitch structure than language and that this is potentially a special feature that makes it unique in human

[215] Ladd, *Intonational Phonology*, 43.
[216] Ladd, *Intonational Phonology*, 44.

communication. But AM theory *cannot account for paradigmatic elements* of pitch patterning in song in which pitch categories are specified on a scale of targets. Since this study posits that such paradigmatic elements are definitive of song melody, and given the essentially reductive formulae employed by the AM framework, it cannot be accommodated into the system developed here.

3.15 Configurations

Another approach to intonational structure—and one that is consistent with the theory developed in this study—is that of the Dutch school.[217] For these theorists *pitch movements* are the basic *perceptual* categories of intonational structure. "These movements—different kinds of rising and falling pitches—form 'configurations' (Prefix, Root, and Suffix) which are in turn concatenated into 'contours'."[218] Unlike models predicated on pitch accents or points in an array, the Dutch school emphasizes the same gradient elements that I have highlighted throughout this study. There are elements of this model that we may adopt in our theory of prosodic structure. For one thing, they allow for comparison of categories according to 'shape' rather than discrete, contrastive features alone. When we study pitch contours we often find that, "[t]he difference between the forms is one of range, which can again be correlated with a semantic parameter. By contrast, the levels analysis cannot relate the contours in terms of common shape, since such shapes do not exist phonologically in this analysis. The relationship

---

[217] The Dutch school of intonation is the product of research conducted at the Institute for Perception Research at Eindhoven (IPO) and is most often associated with the work of three linguists: Antonie Cohen, René Collier and Johan 't Hart (1990).

[218] Fox, *Intonational Phonology*, 295.

thus becomes an arbitrary one."[219] The second benefit has to do with the perennial issue of gradience. Fox points out that "levels are inevitably arbitrary, since the distinction between a narrow rise-fall and a wide rise-fall, for example, is not a discrete one, but a matter of gradience."[220] We cannot measure these gradient categories as discrete elements.

> [T]here would seem therefore to be good reason to adopt a 'configurations' rather than a 'levels' analysis of intonation patterns; the arguments in favor of the former presented above do indeed seem compelling; the recognition of contour *shapes* allows the sorts of generalizations that are required, by excluding the range or extent of the pitch movement. Thus, for example, falling patterns can be grouped together phonologically on the basis of their shape without reference to the height of the start or the extent of the fall. This is not possible if the pitch level of specific points is the only phonological parameter at our disposal for distinguishing different contours."[221]

In our case study of Zulu song prosody in Chapter 4 we find that these shapes have multiple melodic functions that may or may not be associated with categorical features of melodic structure. For instance, a falling glide may signal a phrase boundary rather than the action of a depressor consonant. Shapes such as these cannot be reduced to level tones if they are to be compared as suprasegmental elements. The final section of this chapter suggests how we might conceive of these suprasegmental features within a larger prosodic structure.

---

[219] Fox, *Prosodic Features and Prosodic Structure*, 299.
[220] Fox, *Prosodic Features and Prosodic Structure*, 300.
[221] Ibid.

This discussion of prosodic features in this chapter has focused on tone and intonation because these two features are vital to pitch patterning in many song cultures and are important components of melody. But there is a crucial difference between these two elements that I have not yet addressed in this chapter: this is the issue of arbitrariness. Carlos Gussenhoven (2004) writes that unlike the categories and functions of speech tone, "many linguists have observed that there seems to be something pervasively non-arbitrary about intonation."[1] In both music and language the gradient features of intonation signal emotional and paralinguistic states that are unambiguously meaningful and probably universal. These personal, embodied meanings are conveyed using the same signal that we use for the combination of pitch categories surveyed in this chapter. It is this synthesis of abstract combinatorial and holistic elements that make melody unique as a mode of communication.

3.17 Conclusion

The usage-based model advocated in Chapters 1 and 2 is extended in Part 1 of this chapter where I argued that the concepts of tonality and hierarchical structure should not be tied to conventional categories of discrete pitch. I introduced two new concepts to account for elements of gradience in song melody: pitch targets and glides. These combinatorial units are acquired through a process of social learning based on repetition, or what I describe as chunking. In the second part of the chapter I focused on prosodic features and emphasized the fact that song and speech share elements of pitch accent, tone (the analog to pitch targets and glides) and intonation (the analog to melodic contour). Each of these elements plays an important role in the cognitive architecture of melody, as I will show in the case study of Zulu song prosody in Chapter 4. Pitch patterning in song consists of a wide range of categories that we may conceive as a complex adaptive system. This system is adaptive because it is based on past experience, and past and present experience feeds into future usage. It is complex because fluctuations in pitch simultaneously convey a rich diversity of information as part of a multi-dimensional, networked structure. This information is the product of a combinatorial system that employs a range of abstract pitch categories, but it is also a system that conveys embodied affective information that is meaningful in its own right. To understand how this takes place I have described a number of pitch categories for song and speech that may be fused into a single theoretical framework. This is indeed the rationale for treating melody as prosody: the idea that the gradience and variation we find in melodic pitch categories and structures are shared by song and speech.

CHAPTER 4

Zulu Song Prosody

The relations between song and speech have long been recognized to occupy a special place in the study of African music. The work of musicologists and linguists such as Erich M. von Hornbostel, A.M. Jones, J.H. Kwabena Nketia, John Blacking, Gerhard Kubik, and Kofi Agawu—amongst many others—have detailed the inextricable links between these two communicative systems.[222] One important reason for this persistent interest is the predominance of tone languages in sub-Saharan Africa. Speech tone has been widely recognized as a central, if not determining, factor in song melodies. But the relative importance of other prosodic features, including intonation and accent, has largely been ignored. This is partly due to methods of transcription and analysis that rely on the categories and symbols of staff notation. The inherent properties of its standardized structures are not effective or appropriate for accurate analysis of intonation contours. To model melody as prosody we need take into account gradient features of pitch patterning that should be regarded as categorical in their own right. These

[222] Kwabena Nketia makes the following observation about melodic organization in African music: "Whatever the scale, attention is paid as far as possible to the intonation of the text. This is because distortion of the intonation of phrases or the tones of words might create problems for the listener, for many African languages are 'tone languages,' that is, languages in which tone is phonemic, or serves to distinguish words in much the same way as do vowels and consonants." J.H. Kwabena Nketia, *The Music of Africa* (New York: W.W. Norton, 1974): 184. See also: Erich M. von Hornbostel, "African Negro Music," *Africa: Journal of the International African Institute* 1, no. 1 (1928): 30-62; A.M. Jones, *Studies in African Music* (New York: Oxford University Press, 1959); John Blacking, *Venda Children's Songs: A Study in Ethnomusicological Analysis* (1969, Repr. Chicago, University of Chicago Press, 1995); Gerhard Kubik, *Theory of African Music*, Volume 2 (Chicago: Chicago University Press, 2010); Kofi Agawu, *Representing African Music: Postcolonial Notes, Theories, Positions* (New York: Routledge, 2003).

categories were introduced in Chapter 3 and are the basis for the phonological treatment of Zulu song prosody in this chapter. For descriptive accuracy I use pitch extraction and spectrogram techniques as an alternative to staff notation. Phonetic analysis is widely used in linguistic sciences and in music cognition, and it enables a fluid modeling of pitch patterning in song.

The first part of this chapter surveys pan-stylistic features of Zulu song across regions and dialects. Part 2 focuses on linguistic determinants and the role of tone and intonation in particular, illustrated with selected analyses of recordings from my own fieldwork. Part 3 tests some of the hypotheses set out in the first two parts of the chapter through a case study of ten *memulo* songs, and applies many of the ideas developed in Chapter 3.

PART 1: PAN-STYLISTIC FEATURES OF ZULU SONG

There are a wide range of Zulu song genres that I will survey here. Music and dance are not separate categories in Zulu cultures, even though most genres are identified by dance- rather than song-style. A basic classification of these will prove useful as an introduction to the genres discussed later in the chapter. Zulu ceremonial and dance songs are generally categorized under a single term: *ingoma* (noun: song, dance, hymn). Individual *ingoma* styles are defined in terms of specific regional differences. Gender also plays a crucial role in such classifications since there is usually a clear separation of male and female roles and responsibilities; male and female genres are largely autonomous. Two of the most popular genres of women's dance songs that I recorded are the *memulo* (coming-of-age) and *isigekle* (wedding songs) analyzed in Parts 2 and 3 of this chapter. Men's dance songs include several sub-genres that are regionally distinct and marked by specific choreographies, rhythms, and instrumental accompaniment (or lack thereof). I recorded *indlamu* (central and north-west KwaZulu-Natal) *isishameni* (midlands), *umzansi* (south and Drakensberg regions) and *izingilili* (north-east). Vocal songs not associated with dance are *amahubo* (anthems) and *amakhwaya* (mission-influenced Christian choral songs in four-part harmony). Bow music used to be very popular, especially amongst women, but has gradually been replaced by Western instruments like the guitar, harmonica and concertina. The two neo-traditional genres of Zulu popular music are *maskanda*, a guitar-drum-and-bass style, and *isicathamiya*, the male *a cappella* vocal genre with close-knit harmonies and a distinctive call-and-response heterophonic structure.[223] There are several urban genres that include various forms of mbaqanga, jazz, gospel, house, and kwaito.[224]

There are several features of pitch patterning that are shared across a range of song genres, some of which are linked to the prosodic characteristics of isiZulu.[225] The influence of these 'linguistic' determinants on the structure of song melodies is the main focus of this chapter, and it may well account for the fact that the vast majority of Zulu

---

[223] *Isicathamiya* originated in the urban hostels of Johannesburg and Durban where groups of male migrant-workers would join in choral singing after hours and on weekends. These weekend competitions continue today at hostels like Jeppe and George Goch in Johannesburg. See: Veit Erlmann, *Nightsong: Performance, Power and Practice in South Africa* (Chicago: Chicago University Press, 1990).
[224] Kwaito is a form of electronic house music. See: Gavin Steingo, *After Apartheid: Kwaito Music and the Aesthetics of Freedom* (PhD Dissertation: University of Pennsylvania, 2010).
[225] 'isiZulu' refers to the language and 'Zulu' to the ethnicity (see Chapter 5 for discussion).

song is unaccompanied. Historical and contemporary surveys both point to the popularity of choral music over and above instrumental traditions.

> Perhaps the most distinctive feature of the traditional music of the Zulu, as also of other Bantu-speaking peoples of the Nguni group living in the extreme south-eastern corner of Africa, is the fact their communal music-making is exclusively vocal. It comprises dance-songs in which the performers sing their own dance-music. Drums and percussion ensembles, so common elsewhere in Africa, are entirely lacking (though warriors used to beat on their shields with their weapons, and in recent decades percussion accompaniment has been introduced into young people's recreational dances, under the influence of neighboring peoples). For individual music-making, on the other hand, flutes and musical bows of several different types were very frequently used in earlier times, either for solo playing or for song accompaniment.[226]

Rycroft's surveys field research was conducted between the 1950s and 1970s, and yet the prevalence of dance songs and other *a cappella* vocal numbers is consistent with my own work in rural KwaZulu-Natal. Traditional instruments are increasingly rare. Regardless of this fact, it seems that the structural features apparent in unaccompanied music seem to hold for the neo-traditional popular musics like *maskanda*[227] and *isicathamiya*, too.[228]

---

[226] David Rycroft, "A Royal Account of Music in Zulu Life with Translation, Annotation, and Musical Transcription," *Bulletin of the School of Oriental and African Studies, University of London* 38, No. 2 (1975): 351-402.

[227] "*Maskanda* is a Zulu version of the Afrikaans word for musician, *musikant*. In KwaZulu-Natal what Europeans called "music" was *imusic*. All other forms of Zulu traditional performance were called *ngoma* (song-dance-drumming-healing complex). Clearly *maskanda* was associated with music made by Afrikaans-speaking people, in this case, White farmers who played their music for relaxation on the farms where Zulu men worked in the nineteenth and early twentieth centuries. The word *maskanda* then suggests three levels of musical and cultural encounter in the shaping of twentieth-century Zulu guitar performance: (1) European music, (2) translated into Afrikaans cultural practices, (3) borrowed and transformed by Zulu-speaking musicians into a musical language more consistent with their own aesthetics but played on

Although Rycroft and also Percival Kirby reported on the use of several bow instruments and flutes, these seem largely to have been forgotten.[229] In fieldwork conducted for this project (see Chapter 5) bow instruments were found in just one area (where they had recently been re-introduced by a visiting ethnomusicologist).[230] This is an important point for our study of song melody because musicologists have long explained tonality according to the harmonic principles of the *ugubhu* and *umakhweyane* bows. Today these bow instruments have been replaced by the guitar and concertina as the most popular instruments for personal and group music making. Guitars are commonly found in both rural and urban areas and are used in *maskanda* music. Techniques are novel and widely divergent, and very few players adhere to even the simple chord patterns characteristic of

---

European instruments like violin, concertina, piano accordion, and most important, the steel-string guitar." Carol A. Muller, *Focus: Music of South Africa*, Second Edition (New York: Routledge, 2004): 113-114. Today, most if not all Zulu music are syncretic of both 'African' and 'European' forms. I will not attempt to establish the boundaries between these. The continuities and discontinuities of Zulu musics past and present is discussed in greater detail in Chapter 5. The important point for this discussion is that there is a coherent shared practice specific to a community regardless of its appropriations and assimilations. It could well be argued that syncretism is the very hallmark of twentieth century South African music.

[228] Millions of isiZulu-speakers live in towns and cities but most continue to hold strong links to their rural communities and often return for holidays and important ceremonies, especially for weddings, funerals and *memulos*. This is out of respect for their ancestors as much as for their families, friends and communities as a whole. There are still music practices in urban areas that are distinctively 'Zulu' and which show continuities between urban and rural. These include many of the male dances, including *indlamu*, *isishameni*, and *umzansi*, but also the neo-traditional genres of *maskanda* and *isicathamiya* that originated in the experience of urban-rural migration.

[229] Kirby, after all, conducted his original fieldwork in the 1920s and 1930s and Rycroft was most active in the 1950s and 60s.

[230] Ethnomusicologist Dave Dargie also surveyed Zulu music traditions in northern and eastern KwaZulu-Natal and found no evidence of extant bow music practices (personal comment, East London, July 2013). The bow instruments I recorded were the *umakhweyane* and *umqangasi* (see Kirby 1965 and Joseph 1987 for more detailed information on these). Brother Clement Sithole of the Inkamana Abbey near Vryheid taught me how to build both an *umakhweyane* and an *ughubhu* but also lamented the disappearance of these traditions across most of Zululand. (Percival Kirby, *The musical instruments of the native races of South Africa*, Second Edition (Johannesburg: Witwatersrand University Press, 1965); Rosemary Joseph, "Zulu women's Bow Songs: Ruminations on Love," *Bulletin of the School of Oriental and African Studies, University of London* 50, no. 1, (1987): 90-119.

Western common practice. Given these developments it seems unlikely that the harmonic practices of early bow musics can still be considered as the basis for vocal song.

Despite significant changes in Zulu society through the twentieth century, there seem to be features of song structure that are consistent across genres. These include some quite general features like: call and response, interlocking of voice parts, a cyclic phrase structure with melodic variants, and a linear rather than harmonic conception of melodic structure. More specific pitch elements are introduced by factors of speech tone and intonation. These factors are what account for the fact that in ceremonial, dance, choral, and *isicathamiya* songs *there are no pitched instruments*. The requirements of speech tone seem to minimize the practice of homophonic, harmonized singing. Interlocking polyphony and heterophony are more common. The next section of this chapter reviews theories of Zulu tonality as the basis for our investigation of melodic and prosodic features to follow.

1. TONALITY

Tonality may be defined as the system of pitch relations that order musical practice in a coherent cultural group. In many Zulu cultures, this system is *intrinsic* and is not codified in theory or realized in notation. Singers do not consciously reflect on a store of scales, harmonies, or other pitch rudiments to construct their melodies. Other mnemonic devices are employed in composition and performance. Scholars of Zulu music have tended to emphasize the harmonic basis for tonality. John Blacking, for instance, argued that Nguni music is characterized by "regular shifts of tonality" between chords or harmonic "poles."[231]

> Much Black African music seems to be derived from a conceptual framework of chords rather than single tones. Evidence of this is provided by the principle of "harmonic equivalence": if alterations in melody are precipitated by changes in speech-tone, or if other singers are "filling out" a song with extra lines of melody, the tones must be systematically selected from tones which occur in the "chords" that implicitly accompany each shift of tone in the basic melodic pattern.[232]

According to Blacking, similar harmonic principles operate in polyphonic music. "Sometimes the filling out of parts follows the pattern of a pre-conceived framework in which every tone is conceived harmonically. In other cases chords seem to arise from the overlapped antiphony which can be a consequence of the call/response pattern which is a fundamental form in vocal music and often replicated in solo instrumental music."[233] There are reasons to be skeptical of this 'harmonic' viewpoint. I have already pointed out that instrumental music is secondary to vocal music and that pitched instruments are rare in the singing of dance and ceremonial music.

---

[231] The most extensive survey of southern African music systems was presented in the two entries on 'Indigenous Music' for the *South African Music Encyclopedia* edited in four volumes by Jacques P. Malan (1975-1982). Although this was only a segment in one volume of a work dedicated almost exclusively to the music of white and 'European' South Africa, it nevertheless attracted the attention of major contributors and the result could well have been published as its own volume. Musicologists and ethnomusicologists Percival Kirby, John Blacking and David Rycroft drew on a small but detailed literature from comparative musicology, ethnomusicology and linguistics.

[232] John Blacking, "Indigenous Music," in *South African Music Encyclopedia*, ed. Jacques P. Malan (Pretoria: Human Sciences Research Council, 1982): 295-6.

[233] Blacking, "Indigenous Music," 299.

But there is additional evidence that vocal music is primarily linear in conception. For instance, when a soloist renders a multipart song on her own she tends to sing elements of both the call and response. This demonstrates the integral relation between these parts in the mind of the singer. Given that no single melodic line is complete without its complement it would be very awkward for a singer to reproduce just the call or the response on its own.[234] This suggests that the singer has in mind an interlocked pair of melodic lines neither of which are encoded as the working out of harmonies. A second reason is that when combinations of simultaneous tones do occur in Zulu song they are usually sung as unisons rather than other consonant intervals like thirds or sixths. Taking these facts into account, we may reason that Blacking and Rycroft's analytical categories impose on Zulu music the sorts of harmonic functions expected of Western art music rather than their own indigenous principles.[235] The remainder of this chapter focuses on analyzing the linear elements of pitch patterning in song. I begin with the issues of scale systems and then move on to discussion of the call and response principles so pervasive in all song genres.

---

[234] The linear dimension is present even in the basic alternation of voice parts: "[A] fundamental principle of Zulu and Swazi multi-part vocal music is the non-simultaneous entry of voices. In any choral song, there are at least two voice-parts, singing *non*-identical words. These parts never begin together." David Coplan, *In Township Tonight: South Africa's Black City Music and Theatre*, Second Edition (Chicago: University of Chicago Press, 2007): 40.

[235] This seems to have been common practice for Blacking who often made such parallels. He compared Venda folk songs, for instance, to Beethoven and Mahler symphonies. "Broader tonal and harmonic structures affect their shape no less than in scores of the melodies of European 'art', 'folk' and 'popular' music." Ibid.

Bongani Mthethwa observes that there is a great deal of variation in Zulu scale systems, though the most common are pentatonic and hexatonic.[236] Rycroft is in agreement when he states that there is "considerable diversity, not only between the scale system used by different Nguni peoples, but also within single language communities."[237] This variation extends to the number of tones used (tritonic, tetratonic, pentatonic, hexatonic, heptatonic and octatonic) as well as the range of intervallic structures.

> It may perhaps be broadly stated that perfect fourths and fifths appear to be important structural intervals in a great deal of Nguni music, and that in many of the various scales or modes, larger, unbridged intervals such as a fourth or third tend to lie near the bottom, and closer intervals towards the top. This suggests affinity with partials of the harmonic series, as has often been affirmed by P.R. Kirby. But not all Nguni scales allow this inference to be made. Scale systems are not completely homogeneous among the Zulu and Swazi, nor even within the Zulu-speaking area itself. As previously outlined, this is consistent with history, though the sources of different scales or modes cannot as yet be clearly traced. It seems necessary, however, to distinguish broadly between Swaziland, Zululand, and the rest of Natal—though a far more comprehensive survey of the latter area would be desirable.[238]

Rycroft finds no 'homogenous' system across the Zulu-speaking regions. In his study of 30 songs by Princess Constance Magogo KaDinuZulu he demonstrated a great deal of diversity even within the repertory of a single singer: "One song is tritonic; 11 are

---

[236] Bongani Mthethwa, *Zulu Folksongs: History, Nature and Classroom Potential* (BMUS dissertation: University of Natal, 1979): 2.
[237] David Rycroft, "Indigenous Music," in *South African Music Encylopedia*, ed. Jacques P. Malan (Pretoria: Human Sciences Research Council, 1982): 322.
[238] Ibid.

tetratonic; 14 are pentatonic; 2 are hexatonic; and 1 (the children's nursery jingle) is heptatonic. Item no. 23 has been omitted since fixed pitches are not employed."[239] My analysis of *memulo* songs in this chapter also demonstrates this diversity.[240] As I pointed out in Chapter 3, fixed inventories of pitch targets with tonic attractors act as important ordering devices for melodic structures. But the extraordinary variation we find in Zulu scale systems suggests that there may be other pitch categories that structure melody, and that this structure may be more gradient than scholars have assumed.

2. CALL AND RESPONSE

Zulu songs are polyphonic and usually antiphonal, joining two voice parts, soloist and group, in call and response. The call may be prolonged in its first statement, but it ultimately falls into a regular phase with the response. There is often imitation between the two, but they are seldom identical. Instead an interlocking, heterophonic texture predominates. The call and response sections do not act as a distinct closed units but as overlapping and integrated entities, neither of which is independent. The call is usually at a higher pitch level than the response. The response may add rhythmic impetus or melodic counterpoint, and often serves as an 'affirmation,' or chorus line for the soloist. Let us consider two examples that show this characteristic alternation of voice parts. The first is sung by a group of men from the Mweni district. In this song there is a

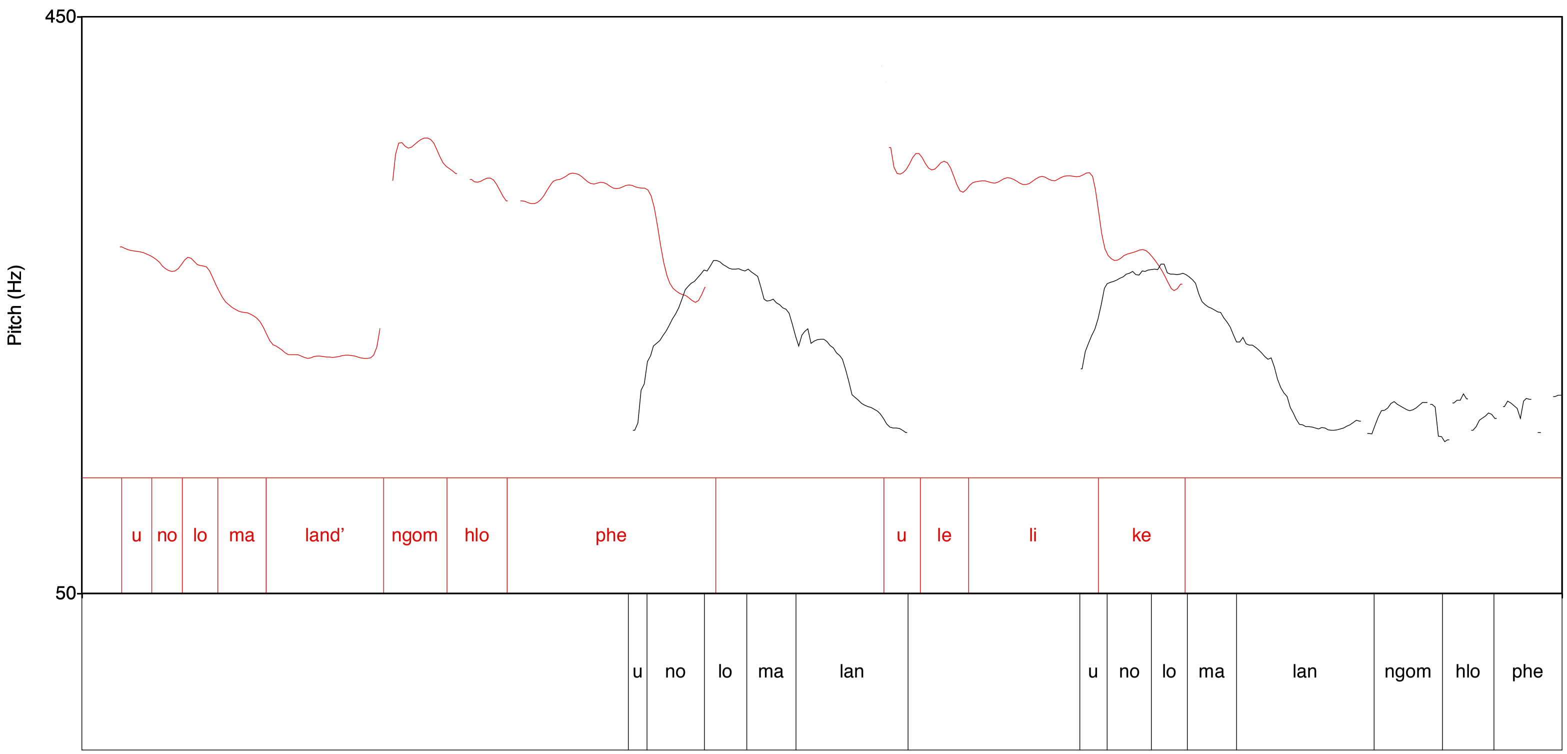
[239] Rycroft, "A Royal Account of Music in Zulu Life," 378.

[240] The problem of non-discrete pitch is an interesting one because it figures in many Zulu songs. The suggestion made in the previous chapter is that we treat such gradient features as contours or as pitch targets. This benefit of this approach is that we retain pitch categories without artificially enforcing a system of level pitch.

characteristic ‘soprano’ solo (the call) and a ‘lower’ choral group response in the bass. The pitch track in Example 2 shows the fundamental frequency of the soloist in red and the choral response in black. The textgrids that figure underneath show the alternation and overlapping of parts and their interdependence. Note that the ‘call’ segments of *uNolomaland’ Emhlophe* are at consistently higher pitch levels. There is always overlapping between call and response lines so that the two are never entirely distinct. Note also how the pitch contours overlap and intersect. The similarities in the pitch contours are also interesting, and point to the speech tone factors inherent in the shared lyrics.

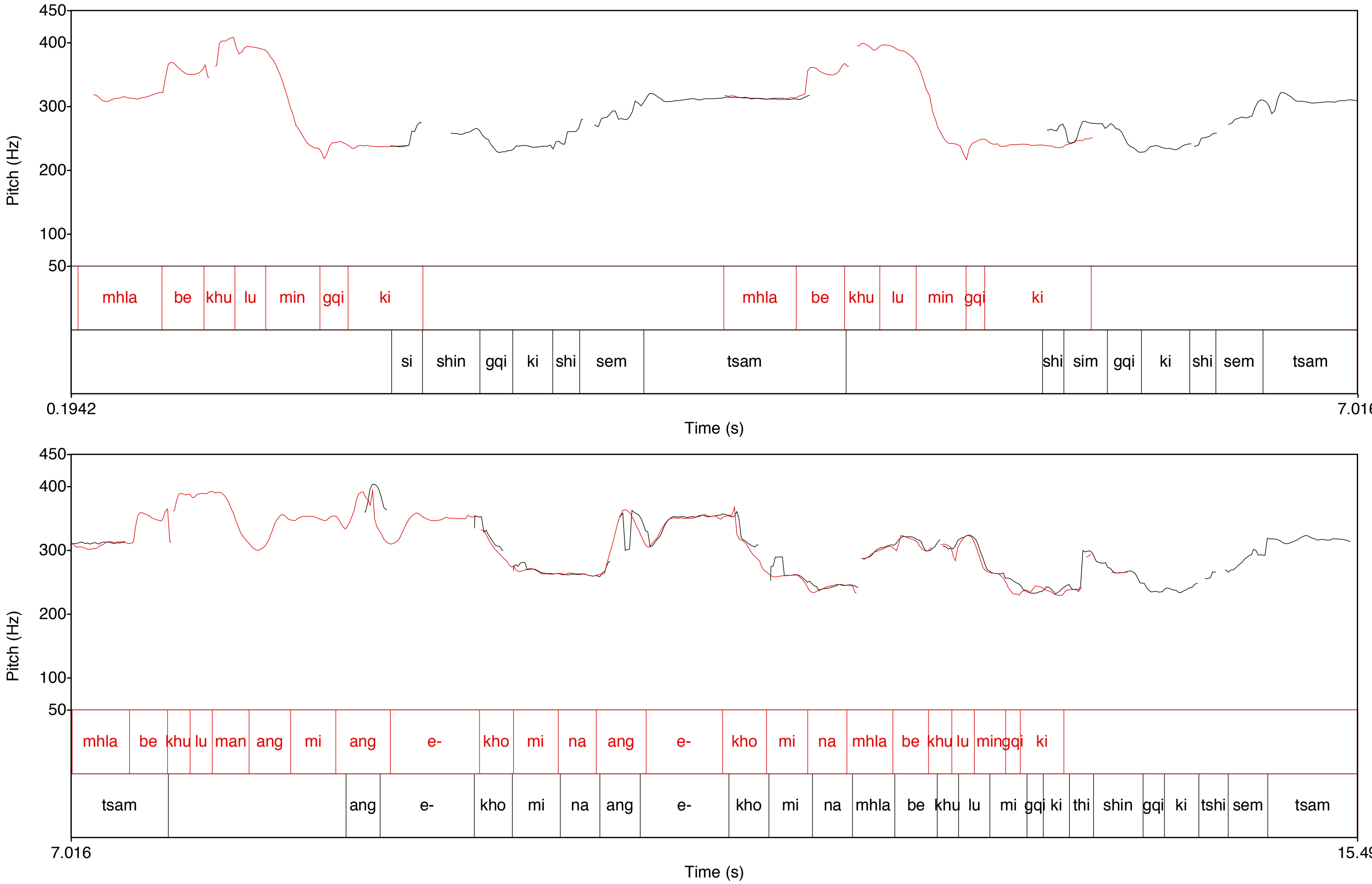
The next example is a girls’ song recorded at Mashunka in the Msinga district. Here the two parts are less distinct in tessitura. The pitch ranges of the two parts overlap and the chorus from ‘angekhe’ is sung in unison with the return of the call. A soloist sings the pitch reduction and text marked in red, whereas the group as a response sings the remainder in black. The initial two entries in call and response overlap. The group response is sustained through the entry of the soloist into the second and third entry, providing a short moment of heterophony. In the second half of the performance the two lines overlap completely where *the same lyrics are sung in both voices*. In Example 2 (*uNolomaland’*) we see overlap between the beginning and ending of phrases because the parts are sung at different pitch heights. Here, in an undifferentiated choir, voices may sing at the same pitch height, and in unison, even in call and response.

Example 2: *uNolomland' Emhlophe* (call and response pattern)



This example shows the characteristic call-and-response pattern of Zulu songs. The overlapping pitch tracks of the call (red line) and response (black line) are depicted in the upper system. Notice that the call is at a consistently higher frequency than the response. The contours of the two trends are similar rather than contrastive. In the lower system we see that despite the similarities in the lyrics, there are important variants that set up a dialogue between the parts.

Example 3: *Mhlabe (*call and response pattern)

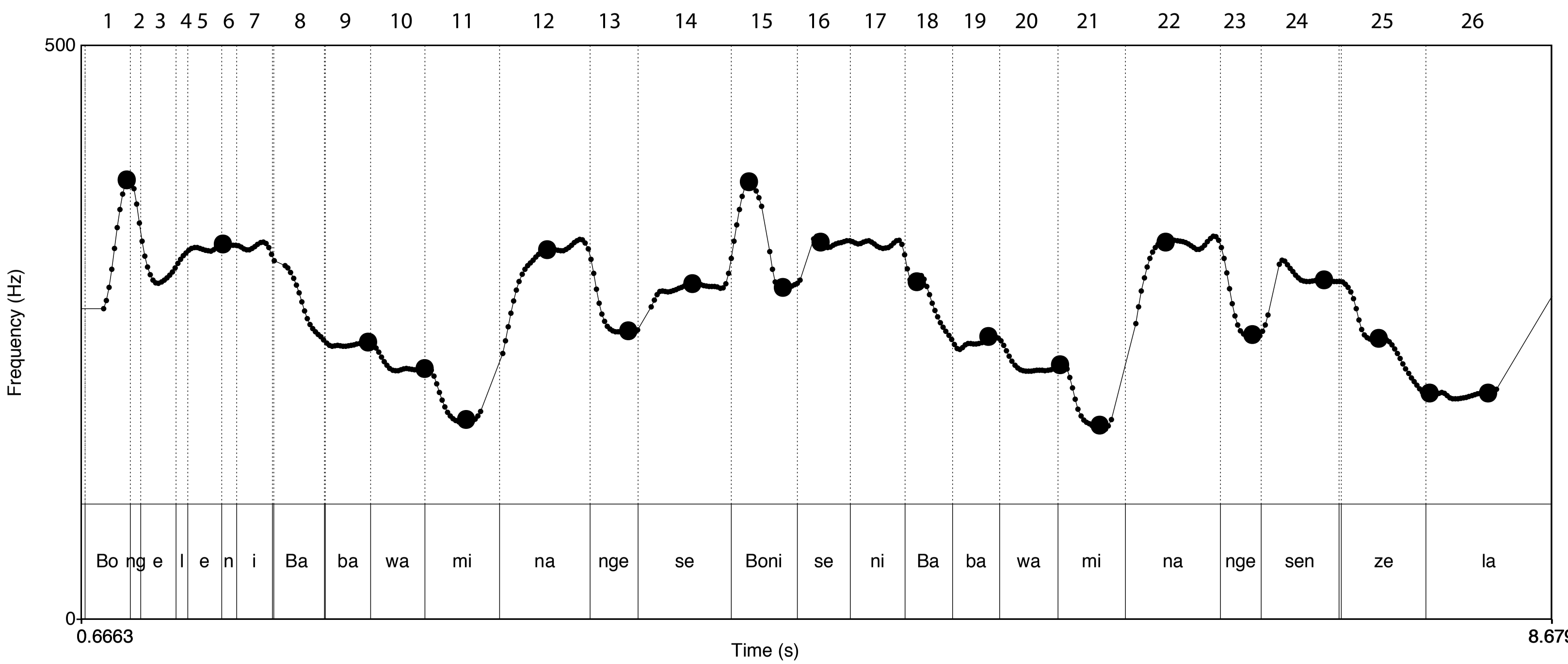


This example shows the close relationship between call and response in the song *Mhlabe*. The red line is the call and the black line the response. Notice how the call begins the song at a higher frequency and that it overlaps entirely with the onset of the response. This song is sung by girls who share the same voice range. The previous song (*uNolomalan*d') included a mixed choir of male voices in which a soprano voice was used as the call.

Examples 4.1 and 4.2 demonstrate, in a third song, that the singers internalize the call and response principle and perform this structure even when singing alone. Were *Baba Wami* sung in a group context, we would see a pitch analysis similar to example 2. In 4.1 I have imposed a series of pitch targets on the fundamental frequency track. Since targets of this sort are inherent and are generally not realized as distinctive pitches in performance, these targets merely indicate tonal attractors to which singers orient themselves. They are cognitive foci. Consistent peaks and troughs in this song show that there is a pitch structure to this performance and pitch movements are not entirely random. Notice that I have not included pitch targets in each and every segment. This is because there is generally only a single target at a particular pitch height.

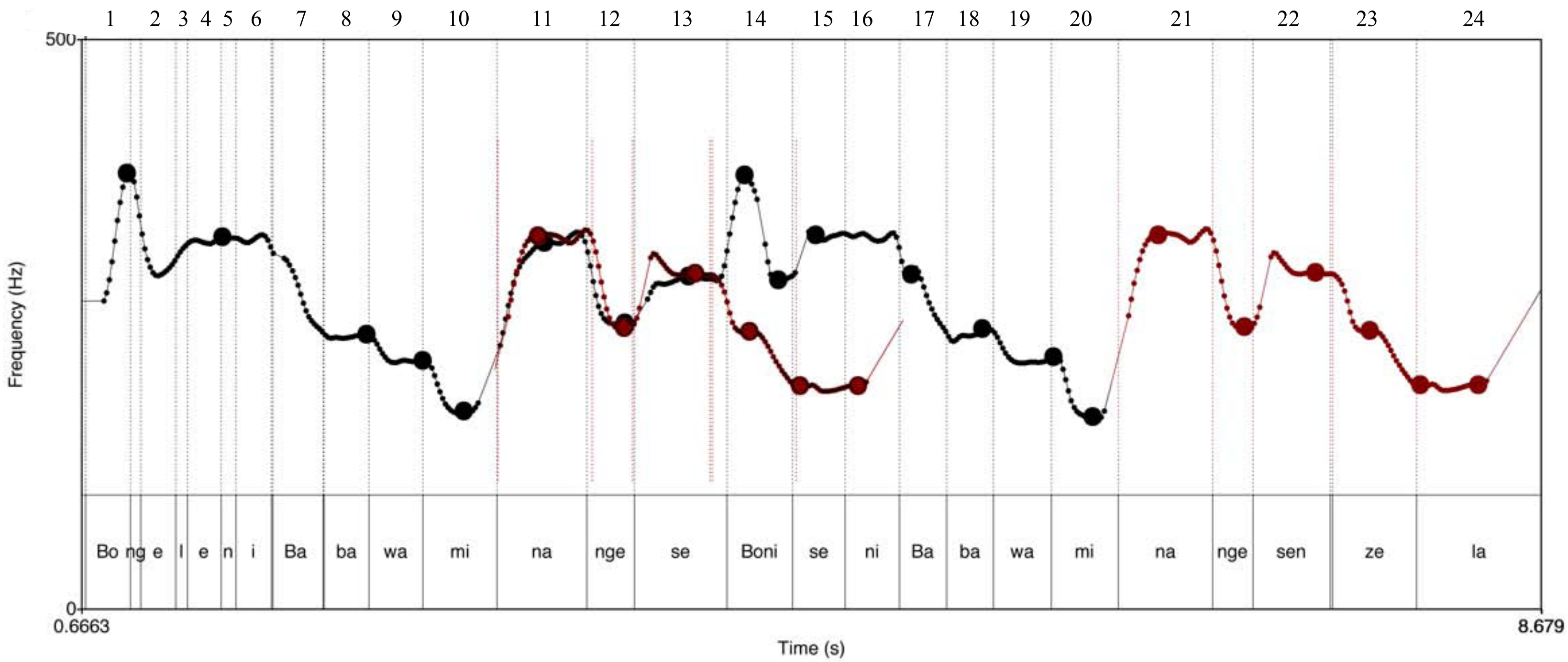
Example 4.2 shows how the singer must incorporate elements of both the call and the response of *Baba Wami* into her melody. This is because the 'call' is short and incomplete without the response. The two form a single Gestalt. However, because there is overlap or heterophony, the singer must cut short the response in order to remain in phase with the next entry of the call. This shows how the singer has internalized the interlocking phase structure. To show this, the second and complete rendition of the responsorial phrase 'nange senzela' (in red) has been superimposed on the partially-realized first entry to show how its full continuation.

Example 4.1: *Baba Wami* pitch targets



This example includes a series of pitch targets that I have imposed on the pitch track. Since targets of this sort are inherent and are generally not realized as distinctive pitches in performance, these targets merely indicate tonal attractors to which singers orient themselves. They are cognitive foci. Consistent peaks and troughs in this song show that there is a pitch structure to this performance and pitch movements are not entirely random. Notice that I have not included pitch targets in each and every segment. This is because there is generally only a single target at a particular pitch height.

Example 4.2: *Baba Wami* pitch targets showing internalized call and response pattern



Example 4.2 is a performance of *Baba Wami* by a soloist. It shows how the singer must incorporate elements of both the call and the response into her melody. This is because the 'call' is short and incomplete without the response. The two form a single Gestalt. However, because there is overlap or heterophony, the singer must cut short the response in order to remain in phase with the next entry of the call. This shows how the singer has internalized the interlocking phase structure. To show this, the second and complete rendition of the responsorial phrase 'nange senzela' (in red) has been superimposed on the partially-realized first entry to show how its full continuation.

3. PHRASE STRUCTURE

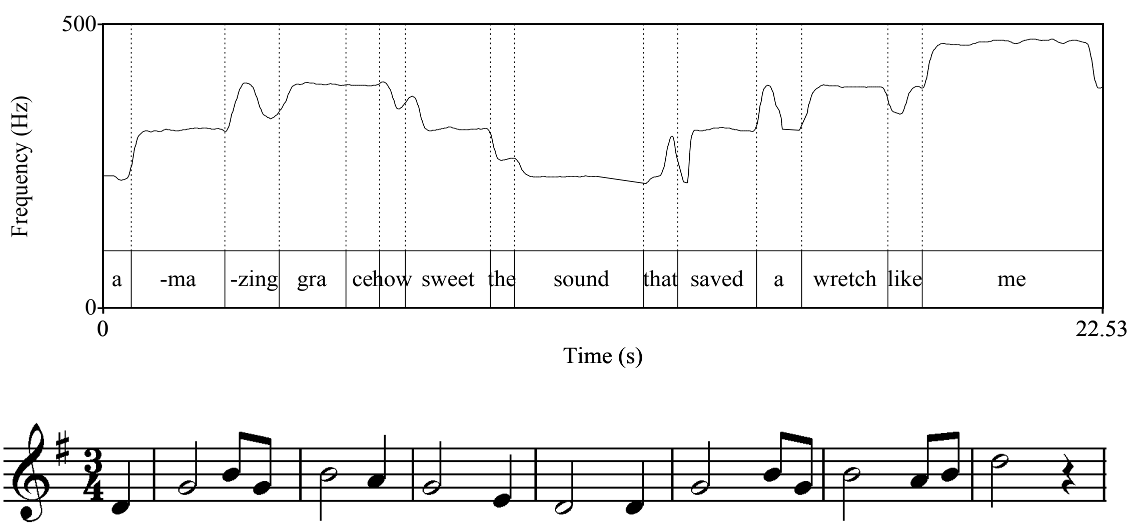
The call and response features of these three examples demonstrate the principle of interlocking so characteristic of Zulu vocal music. There may be just two elements to this interlocking in a simple binary phrase structure or there may be four or even six elements, depending on the complexity of the cycle and on the nature of the 'variants.' I use the term 'phrase' to designate a single and semi-autonomous interlocked call and response unit, whereas a 'cycle' consists of a number of phrases that operate according to an overarching melodic and rhythmic structure. (See the analyses and discussion in Part 3 below.) There is no indication that pitch patterning is structured beyond the cycle since it has no development of melodic material. This may be because of how these songs facilitate and engender social practices that are participatory. Dance songs are performed at weddings, *memulos* and other events to mark rites of passage, to offer commentary and critique, and to celebrate. Songs consisting of short cycles and variants are performed over extended periods of time (ranging from two or three minutes to as many as ten or more) without significant change in the structural units. The duration of a song depends not on the complexity of structure but on how they afford opportunity for group interactions. To some extent these are conditioned by the state of arousal of the performers, dancers, and audience members. Favorite songs will be sung more than once at the same event without a sense of redundancy.

There are usually at least two phrases per cycle. These core motives may be embellished or re-ordered as variants, but ultimately the phrase structure remains regular and there are no structural breaks between cycles. Cadences are rare even if there are certain melodic and durational markers that indicate points of repose and *relative closure*. In practice, songs do not cadence to endings but simply fade out as participation dies down or another singer interjects a new lyric for the next song. Even where a singer-dancer gesticulates and breaks out into a passage of free declamation or *izibongo* (praises) the group continues unhindered. A soloist may not be the author of the 'call' but one of the other singers or dancers who feels the inspiration or occasion to extend him/herself. This continuity ensures that the antics of the soloist do not fall flat and establishes a space conducive for such display. The participation of the group response creates a continued rhythmic and melodic ostinato that is sustained regardless of what is offered by way of interjection in the 'solo' parts (i.e. note the distinction between 'call' and 'solo'). Most choral dance songs are multilayered in performance. That is, many other singers may contribute variants or gesticulations as commentaries on the songs without these fitting precisely to the cycles. This multi-layered text embodies the principle of music and dance as social occasions that afford opportunities for display. The call and response and cyclic characteristics of choral dance songs have several important tonal consequences too. For one thing, they eliminate the sorts of 'complex collective cadences' indicative of Western tonal music idioms. Hierarchical structure is local rather than global and there are no functional harmonic elements evident. Indeed, nothing is 'formalized' and tonal features appear to be in constant flux. It is important, therefore, that we investigate the role of 'discrete pitch' in Zulu song.

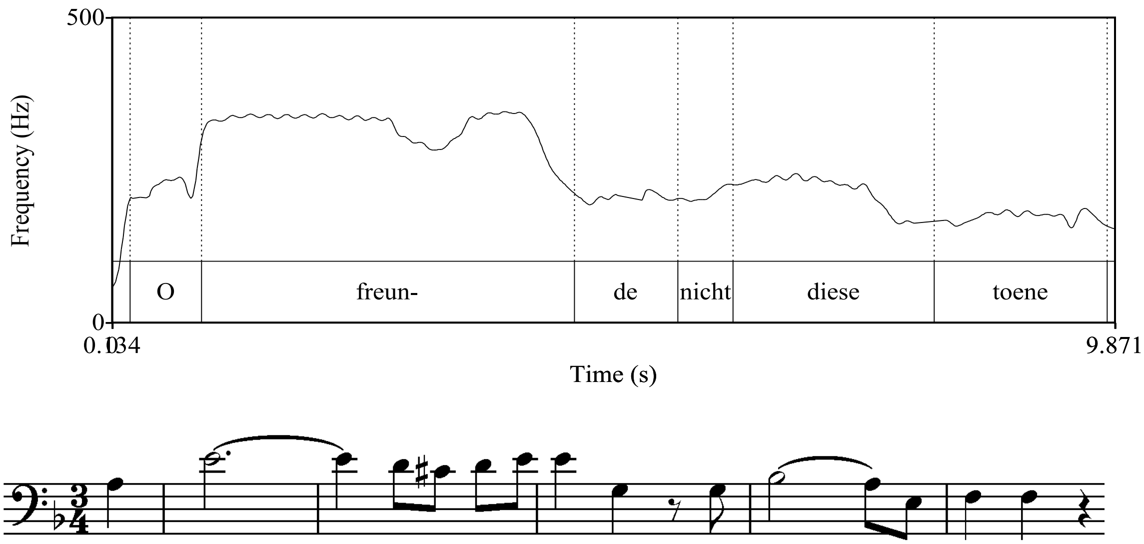
4. PITCH CATEGORIES

Any system of tonality relies on the organization of pitch categories according to an underlying hierarchical structure. These categories are conventionally understood to consist of *discrete pitches* rather than gradient categories. But we have seen that such discrete categories are not appropriate for Zulu songs that show considerable gradience in pitch patterning. Consider the following examples that compare two Western tonal music performances with the Zulu song *Makoti*. Example 5.1 illustrates a pitch reduction from a performance of *Amazing Grace* by LeAnn Rimes. Example 5.2 shows an excerpt from the final movement of Ludwig van Beethoven's Ninth Symphony. In both of these, the pitch perturbations and levels demonstrate the very close relationship of discrete pitch to staff notation and the sorts of pitch features this imposes on the performer. Compare these to the Zulu example of *Makoti* in 5.3 below.

Example 5.1: John Newton, *Amazing Grace* (1779) as sung by LeAnn Rimes.



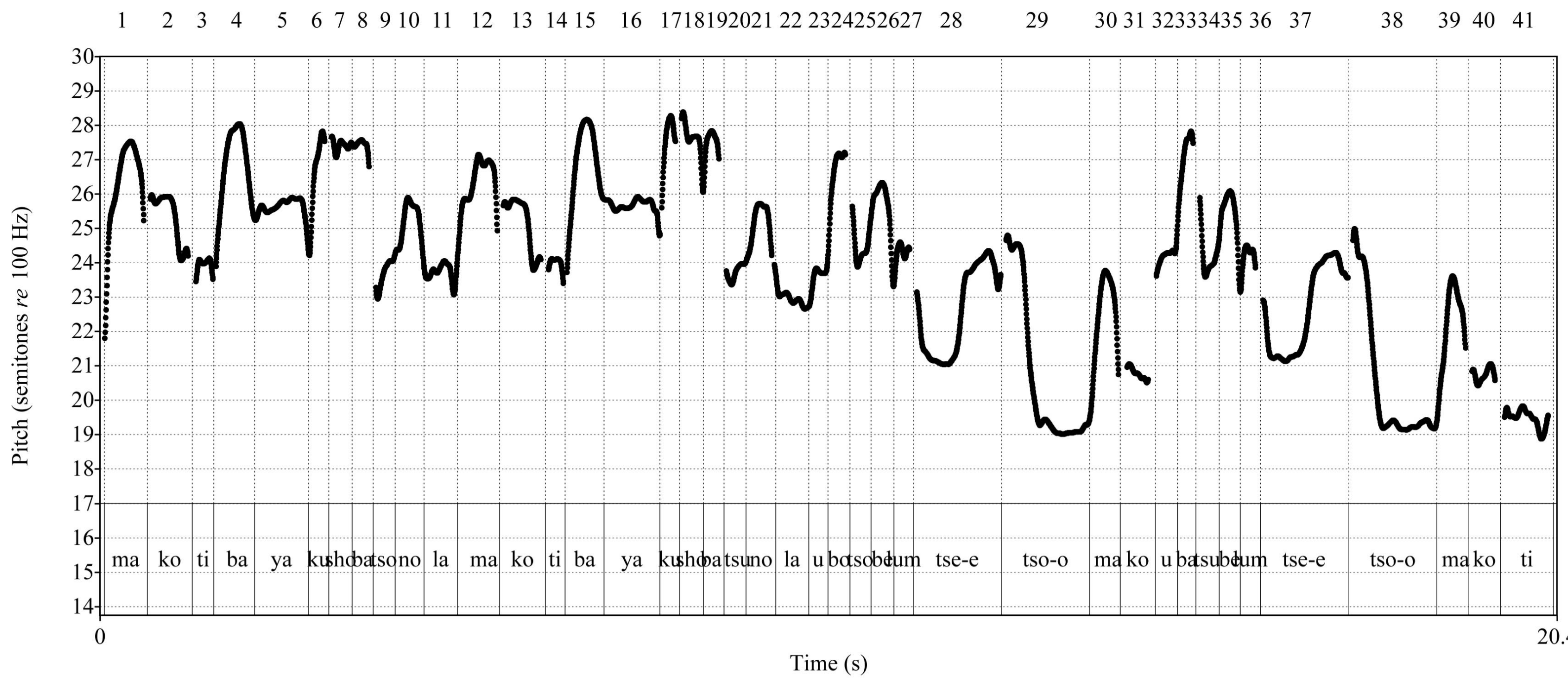
Example 5.2: Ludwig van Beethoven, Symphony No. 9 in D minor, Opus 125, V, Presto, Recitative: “O Freunde, nicht diese Töne!” (O friends, not these tones!)



This extract from Beethoven’s Ninth Symphony is the opening ‘recitative’ passage and yet, like the previous example, shows how singers target and approximate the level pitch values of the score. There are embellishments, but these are minor additions.[241] The point of these two unaccompanied examples is to show how—even in the absence of instruments—singers aim to reproduce discrete level-pitches. They deliberately approximate pitch categories using learned schemata of which intervallic configurations are basic components. No singer could learn this melody without a clear sense of how to reproduce intervals from a semitone to a tone and all other divisions of the octave. This store of learned schemata is not evident in many Zulu songs. The wedding song, *Makoti*, depicted in Example 5 shows how pitch categories seldom occupy the level pitch categories we find in much Western art music.

[241] The initial slide on ‘O’ is explained by it being the first entry; the soloist is likely searching for the note.

Example 5.3: *Makoti* with semitones marked to show gradience



The many peaks, troughs and gradients evident in this example demonstrate the gradience of pitch patterning in Zulu song. The semitone scale on the Y-axis shows the extent to which graded semitone categories are elided, not only by minor perturbations, but also by glides and contours. Of course, these continuously varying 'visual' facts should not divert us from a cognitive reality that is nevertheless categorical: there are *pitch targets* in operation regardless of the duration or discreteness of their realization as level pitches (as shown in examples 4.1 and 4.2 above).

I have included only the opening phrases of *Makoti* because the song is cyclic. Unlike the *Amazing Grace* and "O Freude" examples, there are few instances of 'level' or 'discrete' pitch, as we commonly understand them. The many peaks, troughs and gradients evident in this example demonstrate different sorts of categories. The semitone scale on the Y-axis shows the extent to which classic semitone categories are elided not only by minor perturbations but also by glides and contours. These continuously varying 'visual' facts should not divert us from a cognitive reality that is nevertheless categorical: there are *pitch targets* in operation regardless of the duration or discreteness of their realization as level pitches. What is unusual about Western art music is the extent to which the categories of discrete pitch are standardized and sustained and how they approximate the notational conventions of much instrumental music. Gradient categories, on the other hand, are ignored. To make sense of non-Western, non-discrete pitch contours we need an alternative conceptualization of tonality itself. We need to shift from the analysis of 'discrete pitch,' modeled as it is on common practice tonal music and the categories of staff notation, toward a model that incorporates gradient features like 'pitch targets' and 'contours.'

The best way to model these features and to recognize their significance is through phonetic analysis. Pitch patterning in speech uses many of the same categories as pitch patterning in song, and the instrumental methods designed for studying the physical properties of the human voice are very useful for studying song prosody. In the next two sections I discuss the pitch categories that comprise prosodic structure in Zulu song.

PART 2: PROSODIC FEATURES

General prosodic features were introduced in Chapter 3 and so here I focus only on elements of tone and intonation in isiZulu. These two clusters of elements are important to the study of melody in many African cultures, as J.H. Kwabena Nketia pointed out in his classic survey of African musics:

> When texts in tone languages are sung, the tones used normally in speech are reflected in the contour of the melody. Thus, melodic progressions within a phrase are determined partly by intonation contour, and partly by musical considerations. Sequences of repeated tones and the use of rising and falling intervals or of flexures (fall-rise and rise-fall patterns) in melodies may reflect the intonation patterns used in speech.[242]

These "rising" and "falling intervals or flexures" are commonplace in the tone systems of Bantu languages and they play an important role in melody. As we saw in example 4.3, pitch contours in song cannot be reduced to a series of discrete pitches. The segmentation of a continuously varying pitch contour thus presents a theoretical problem:

> In Bantu languages, spoken utterances display a great many variations of pitch. However, just as in English, there is no fixed scale of absolute pitch values like that expected in music. The multiplicity of pitches used in speech has often made it difficult for investigators to isolate essential, meaningful tonal contrasts from the total contour. For Zulu, for example, Doke in 1926

---

[242] Nketia, *The Music of Africa*, 186.

> postulated nine different levels of pitch, which he numbered 1 to 9 in descending order when marking the tone patterns of words.[243]

In modern linguistics speech tone has been understood as a contrastive feature based on relative rather than absolute pitch. Doke's analysis is therefore interesting because it suggests a more complex paradigmatic dimension than the conventional High vs. Low of the autosegmental metrical treatment now in common use.[244] It also demonstrates a sophisticated syntagmatic dimension that can be used to analyze the contour shapes of words in terms of pitch targets.

Most contemporary phonologists posit two contrasting levels of pitch for isiZulu: High tone vs. Low tone. This is the standard classification for most Bantu languages.[245] In addition to these contrastive tones, Rycroft recognized gliding pitch in three forms: (1) as on-glides (up-glides) to High tones; (2) down-glides as the result of depressor consonants on High tones; and (3) as glides over a single syllable. I incorporate these three tonal categories into the set of categories described in the case study below. The realization of these pitch categories must however be understood in relation to the intonational features of isiZulu:

---

[243] Rycroft, "A Royal Account of Music in Zulu Life," 304.

[244] Doke's pitch levels were comprehensive and are still recorded in the complete isiZulu dictionary he helped to establish. See: Clement M. Doke, D.M. Malcolm, J.M.A Sikakana, and B.W. Vilakazi, *English-Zulu, Zulu-English Dictionary* (Johannesburg: Witwatersrand University Press, 2008).

[245] See Moira Yip, *Tone* (Cambridge: Cambridge University Press, 2002): Chapter 6: African languages.

> In tone languages there is usually no correlation between high pitch and intensity stress. The various syllables of a given word must rise or fall, relative to each other, in a particular way, though alternative patterns may be required in different contexts. However, besides these inherent word patterns there is also an overall "carrier wave" of "sentence intonation". In most Bantu languages this follows a gradually descending contour for normal statements. The patterns of individual words in the sentence cause momentary higher and lower deflections in the descending intonation contour, from syllable to syllable; successive high and low tones, respectively, take progressively lower actual pitch, especially towards the end. In consequence, a high tone at the end may come to be rendered at equal, or even lower actual pitch than an early low tone. Questions, as against statements, usually take a different overall intonation contour. In some Bantu languages this amounts to suspension of a progressive "downdrift", so that successive high tones take equal pitch, instead. The general multiplicity of speech pitches thus appears to result from at least two interacting factors in such languages: (i) an overall "carrier wave" of intonation, which may vary to convey different nuances; this might be viewed as "intermodulating" with (ii) the "essential" tone-patterns of the constituent words. [246]

Sentence intonation is characterized by a "gradually descending contour for normal statements" that results in "progressively lower actual pitch" in the manifestation of speech tones. Given that contrasts are relative rather than absolute, the functional properties of speech tone are not affected. It does, however, change the shape of the melodic contour and may therefore alter the realization of pitch targets in songs. The ways in which downdrift condenses the pitch space has important effects on the realization of tonal features at the segment, phrase, and cycle levels.

---

[246] Rycroft, "Indigenous Music," 304.

Another gradient element to speech tone in isiZulu is the tone lowering effect of 'depressor' consonants. These include: "all voiced fricatives, clicks, and plosives (except for the implosive *b*), and all compounds containing these sounds. High-toned syllables beginning with such consonants commence with a brief rising up-glide. Likewise, descending down-glides from high tones are conditioned by a succeeding consonant of this type."[247] The depressor function of these consonants, operating in conjunction with downdrift, both result in a distinctive and consistent lowering of the melodic line. This may be used to articulate structural properties like closure or to stress a particular contrast. For instance, Rycroft describes elements of 'stylized *portamento*' in 'dual-tone' glides. "A descending glide of a 4th or 5th quite frequently occurs on the last syllable of a phrase, often serving as a compromise between word-tone and melodic requirements."[248]

To conclude: the linguistic pitch features of speech tone (including depressor consonants) and intonation (downdrift) play a major role in melodic structure. In the next section I discuss phonetic methods for studying these features using spectrographic analysis and fundamental frequency tracking in Praat. This introduces the case study of ten *memulo* songs in Part 3 where detailed analysis of each of these phenomena provides empirical support for these general observations.

---

[247] Rycroft, "Indigenous Music," 306.
[248] Ibid.

5. METHOD

This study relies entirely on field recordings that I made over a period of nearly two years from December 2011 to September 2013.[249] Songs recorded at events and performances provided a context for controlled recordings that I made afterwards and which are analyzed in this chapter. These were made with the purposes of phonetic analysis in mind. Since group singing is the norm I used a multi-track recording technique that isolated individuals and groups. However, for each of the *memulo* songs discussed in this case study I recorded singers individually using vocal condenser microphones with a flat frequency response.

The phonetic analysis of these recordings was conducted in Praat because it provides a clear spectrographic representation of the audio file, includes useful annotation tools and employs a reliable fundamental frequency-tracking algorithm. Pitch is extracted and graphically represented as a continuously varying contour. Where sampling errors occur these can be analyzed and remedied by hand. Unlike staff notation, pitch-tracking software does not segment melody into pre-conceived pitch categories and microprosodic fluctuations in the signal are retained. Praat is widely used by phoneticians and includes comprehensive features for extraction, measurement, analysis, and recording.[250] In my analyses I include textgrids (transcriptions of the words and phonemes) for each of the

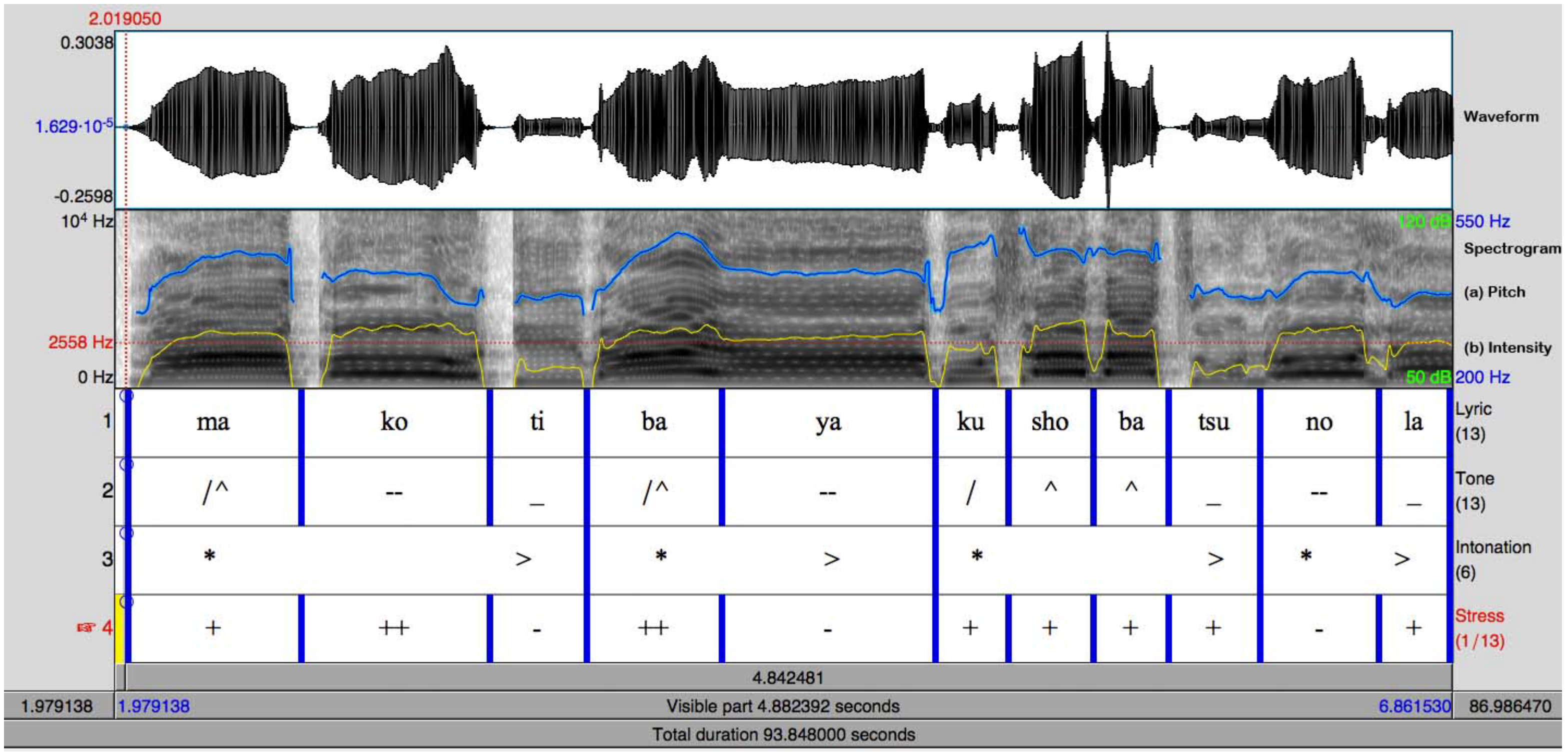
[249] See Chapter 5 for more details on the fieldwork component. Appendixes provide details of the audio and video recordings.
[250] Praat is free computer software for the phonetic analysis of speech sounds that is widely used in linguistics. It includes many different applications for speech analysis, synthesis, statistics, and experimentation. It was designed by Paul Boersma and David Weenink at the University of Amsterdam.

examples. These methods are not unique to Praat, however, as several recent publications show. But Praat is designed specially for speech analysis research and studies in phonetics and phonology.[251] These sorts of methods have been used extensively in studies of music and language.[252] Example 6 demonstrates the utility of this method and also shows how the various tonal features of Zulu song prosody are studied. This screenshot shows the waveform, spectrogram (with pitch and intensity analyses overlaid) and four annotated tiers for lyrics, tone, intonation and stress.

[251] For a review of speech analysis literature see Charles W. Hardcastle, J. Laver and F. Gibbon, eds., *Handbook of the Phonetic Sciences*, Second Edition (Oxford: Wiley-Blackwell, 2010). For music-specific applications of spectrogram techniques see Eric Clarke and Nicholas Cook, eds., *Empirical Musicology* (Oxford: Oxford University Press, 2004). For a historical review see: Cooper, David and Ian Sapiro, "Ethnomusicology in the Laboratory: From the Tonometer to the Digital Melograph," *Ethnomusicology Forum* 15, No. 2 (2006): 301-313.

[252] David Huron and Anirrudh Patel and their collaborators have published extensively in this area. See, for instance: David Huron, "What is Melodic Accent? Converging Evidence from Musical Practice," *Music Perception* 13, No. 4 (1996): 489-516; David Huron, "Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles," *Music Perception* 19, No. 1 (2004): 1-64; David Huron and Joy Ollen, "Agogic contrast in French and English Themes: Further Support for Patel and Daniele," *Music Perception* 21, No. 2 (2003): 267-271; Aniruddh D. Patel, J.R. Iverson and J.C. Rosenberg, "Comparing the Rhythm and Melody of Speech and Music: The Case of British English and French," *JASA* 119 (2006): 3034-3047; and for a review see Patel, *Music, Language, and the Brain*, Chapter 3 'Melody.'

Example 6: *Makoti* analysis in Praat



This screenshot shows an analysis of the first phrase of the song *Makoti* in Praat. Included, from top to bottom, are: waveform (a visual representation of the audio file); spectrogram (visual representation of the sound spectra); pitch track (blue line showing the fundamental frequency track measured in hertz); intensity (yellow line showing the intensity contour measured in decibels); lyric (transcription of the words of the song segmented according to the pitch movements shown in the pitch track); tone (analysis of the pitch categories or speech tones using a standard inventory of glides, high, mid, and low categories); intonation (analysis of the pitch contour showing phrase boundaries); stress (analysis of contrasts in intensity stress).

Several aspects of this short song segment are shown in this composite graph:[253]

(i) Waveform: a visual representation of the audio file from which this example is derived.

(ii) Spectrogram: the representation of the sound spectra and formants [i.e. the resonant frequencies of the vocal tract]. This is the most important tool for analyzing linguistic features. The relative darkness of the patterns illustrates the intensity of the formants. Analysis of formant transitions is crucial for segmenting phonemes. The y-axis (measured in hertz) samples the range 0 Hz to 10000 Hz. Two separate aspects of the song are superimposed on this background:

(a) Pitch track: this blue line depicts the overall intonation contour by tracking fundamental frequency. Fundamental frequency (F_0) is not equivalent to pitch, but it represents an accurate modeling of pitch perception in humans. The F_0 is equivalent, in other words, to the melodic line. This is the most important element for the analyses presented in this study. For each of the case studies I have relied entirely on the analysis of pitch tracks and have excluded other parameters from consideration. A more complex theory of song prosody is beyond the scope of this study.

---

[253] This song was recorded at the Dladla *umuzi* (homestead) in the Msinga district of KwaZulu-Natal on 12 April 2012. It is a wedding song sung by a female soloist and chorus in call-and-response. Extracted here is the 'call' sung by the soloist. The recording was not made at a wedding but in a hut using three different condenser microphones to ensure that soloist and chorus were recorded independently. This method of recording and analysis applies to the other recordings discussed in this chapter.

(b) Intensity track: This yellow line depicts intensity curve based on decibel (dB) level. Here it is confined to the range 50 dB to 120 dB.

(1) Lyric: transcription of phonemes. This is a rough transcription of the isiZulu words and syllables that comprise the lyric to the song. These are segmented according to the formant transitions displayed in the spectrogram.

(2) Tone: I have transcribed individual tones according the following inventory:

/	up-glide
^	high
-	mid
_	low

Although isiZulu tonal features are commonly studied in terms of High vs. Low contrasts, the analyses that I discuss in Part 3 suggest that mid tones are also present in a phonetic description.

(3) Intonation: transcription of intonation segments showing phrase boundaries and pitch movement from High (*) to Low (>). The segmentation of intonational structures is a complex theoretical issue that was raised in the previous chapter. This analysis segments the contour into individual phrase units.

(4) Stress: transcription of emphasized syllables. The symbols used imply:

++	accented

+ stressed

- unstressed

I have included a range of stressed and unstressed (+/-) syllables that correspond with the intensity contour depicted in yellow (b). Stress may or may not correspond to the speech tone. We see this on the second stressed syllable ‘ko’ which I have analyzed as a mid-tone.

This short extract from he song *Makoti* demonstrates several important features of isiZulu tonal phonology:

(i) *High* and *Low* tone contrasts. Example: between [bâ] [tsŭ]

(ii) *Up-glides* for High tones. Example: over the phoneme [má].

(iii) *Down-glides* for depressor consonants. Example: [kò]

(iv) *Glides* or *contour tones* over a single syllable. Example: [bâ]

The realization of these pitch features is relative, not absolute, and is generally bound to the speech tone and articulatory requirements of word segments. The intonation of this short phrase also reveals the phenomenon of ‘downdrift’: a gradual descending contour over the course of the utterance. This is visible in *descents* both within intonation phrases—marked from [*] downward [>], and in the entirety of the pitch track from a high point on [mâ] to a low tone on [lą]. Notice also how features of tone, intonation, and duration mutually reinforce one another to articulate a prosodic structure. Although these parameters seem to be conformant in this example, there are many cases where the interaction is more complexly ordered and for different communicative purposes. Staff notation cannot account for these various prosodic features. Gradient elements such as contour tones, glides, and downdrift are not reducible to discrete pitches. It is also not possible to accurately represent intensity or timbre, both of which are important for the analysis of stress patterns and articulation. In the case study that follows I have used the spectrogram in this way to analyze pitch prosody and to arrive at an accurate transcription and segmentation of the pitch categories critical to the study of melody.

PART 3: TEN MEMULO SONGS

This case study of ten *memulo* songs presents a detailed analysis of the elements of Zulu song prosody surveyed in the first two parts of this chapter. The analyses presented here are based on three sets of recordings and fieldwork studies. I attended six *memulo* ceremonies in four regions of KwaZulu-Natal (Msinga-Weenen in the districts of Nkaseni, Ncunjane, and Mashunka; Ingwavuma at Ndumo; Eshowe at Dakeni; Bergville at Mweni. For more details consult Appendixes A and B),[254] as well as the annual

[254] The duration of a *memulo* may be a week or more and culminates in one or two days of celebration in honor of a young woman or women. It is an ad hoc rite of passage for those between the ages of 17 and 25 (sometimes older), and it registers that preparedness for marriage. A senior girl imparts the values and customs of married life to the initiates and guides them through a period of seclusion. This is followed by a public celebration that is viewed as an honor bestowed upon a young woman by her father. The community and family provide gifts of household items as well as money. Dancing, singing and feasting usually take place over two days in which the initiates and their younger kin take center stage. Groups of young men also take part and compete in choreographed troupes and sometimes as soloists. These proceedings are lead by an elder who may also join in the singing and dancing. The *memulo* must be sanctioned by the *induna* (headman) of the *isigodi* (district) and usually also by the *inkosi* (king) of the tribe to whom a levy must be paid. The success of the event is dependent on the participation of the community, usually including potential or confirmed suitors. Families go to great lengths to prepare food and refreshments at considerable expense. A sacrifice must be made to honor the ancestors, usually of a cow for each of the girls celebrated in the *memulo* as well as goats. Songs tend to describe the challenges, anxieties, and experiences of adolescence and early womanhood. These provide a sanctioned space for topics that may

Umkhosi Womhlanga (Reed Dance) convened by King Goodwill Zwelethini at Nongoma where *memulo* songs are sung.[255] I recorded audio and video at each event and conducted interviews. I also made recordings of several groups of girls singing these songs on two separate occasions and in a controlled environment. This was done in order to eliminate the sounds of dancing, clapping, and environmental noise that obscure and interfere with accurate pitch extraction methods. I then recorded solo and group parts separately using techniques of microphone placement. On my final trip, having analyzed the initial recordings, I set out to record each of the three girls speaking and then singing the songs on their own. Each song analysis relies on these three performances, though not all of them are replicated here in the analysis of each song. For consistency and comparison I have transcribed sung and spoken pitch tracks (see examples 7.1 and 7.2). The different performances reveal important commonalities and discrepancies between each mode of delivery that point to underlying cognitive features that are discussed at the end of the chapter. Call and response elements are discussed where appropriate, and various other graphic analyses are used to supplement these solo records. All of this enables a more or less comprehensive study of pitch relations in these Zulu songs.

---

otherwise be considered taboo. Lyrics are nevertheless obscure to many listeners as I found out in my interviews.

[255] The *Umkhosi Womhlanga* (Reed Dance) is an annual event that celebrates young women virgins. In the past the Zulu king would choose a wife from the assembled company, but today it is a closely monitored educational institution funded in part by the provincial government. Tens of thousands of young women attend and are organized into regional sections and subsections. They sing, dance, and march their way into the Royal Palace where they present the king with a reed collected several kilometers away. The songs they sing are often similar to those used in *memulos*.

Three sets of graphs are provided for each song of the ten songs: The first graph is the song analysis, the second graph is the speech analysis, and the third graph aligns the pitch categories derived from the first two to show the relationship of song and speech categories to one another.

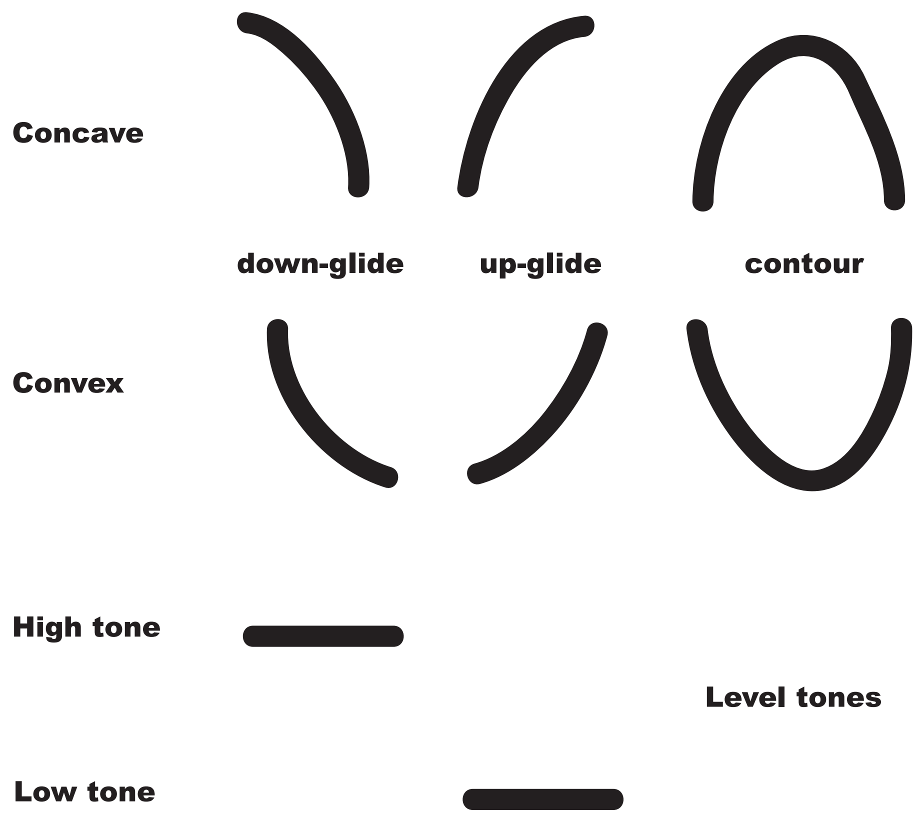
The song graph includes four main elements: The *upper system* in 7.1 shows a pitch track measured in hertz on the Y-axis. This was generated using the graphing functions in Praat. The *lower system* comprises three elements. The first is the textgrid generated in Praat, in which the words of the song are broken down into segments *according to their pitch targets and articulation*. The cross hairs of this grid divide the pitch reduction into segments numbered on the X-axis at the top. Segmentation of the pitch contour is not an arbitrary exercise because it serves to define the pitch categories and their relationship to phonemes and morphemes. We see, for instance, that pitch categories do not always correspond to the syllable segments; the relationship between the two is complexly ordered and is the product of the *suprasegmental* pitch structures discussed in Chapter 3. In some cases there may be two or more pitch targets specified for a single syllable segment such as is the case for glides. (Usually these pitch fluctuations can be explained by physical properties of articulation.)

The two rungs at the bottom of the graph were generated in Adobe Illustrator and Adobe Photoshop. Pitch targets are indicated on the lowest rung. Although one might think that the ruled lines represent a precise intervallic structure, this is not a traditional score.

There are two main purposes: the first is to show the number of pitches used in the melody (i.e. the scale); and the second is to give a sense of the relative heights of pitch targets and the consistency of their realization through an utterance. We shall see that the absolute distances between these pitch targets shows some gradience and does not reduce to a conventional interval structure based on functional relations in a hierarchy. The pitch targets should not be understood as discrete pitches but as categories, many of which are not realized as level pitch in any case. Instead of measuring each and every frequency on the contour we are able to chart its general features, including its contour, based on these pitch targets. They are abstract phonological categories and do not correspond directly to their phonetic representation in upper system of the graph. The pitch targets may also be used to identify pitch movement and to show cadential movement and global structure.

The second-lowest rung of the graph is the symbol set for the analysis of pitch movements over syllable segments. These symbols are used to indicate the categories of speech tone, including the action of depressor consonants within single or conjoined segments. By comparing the individual symbols derived from the *song* graph to those of the second *spoken* graph we have a clear method of comparing the tonal features of song and speech. These symbols may be reduced to four main types: (on- and off-) glides, contours, and level tones.

Figure 1: Pitch categories



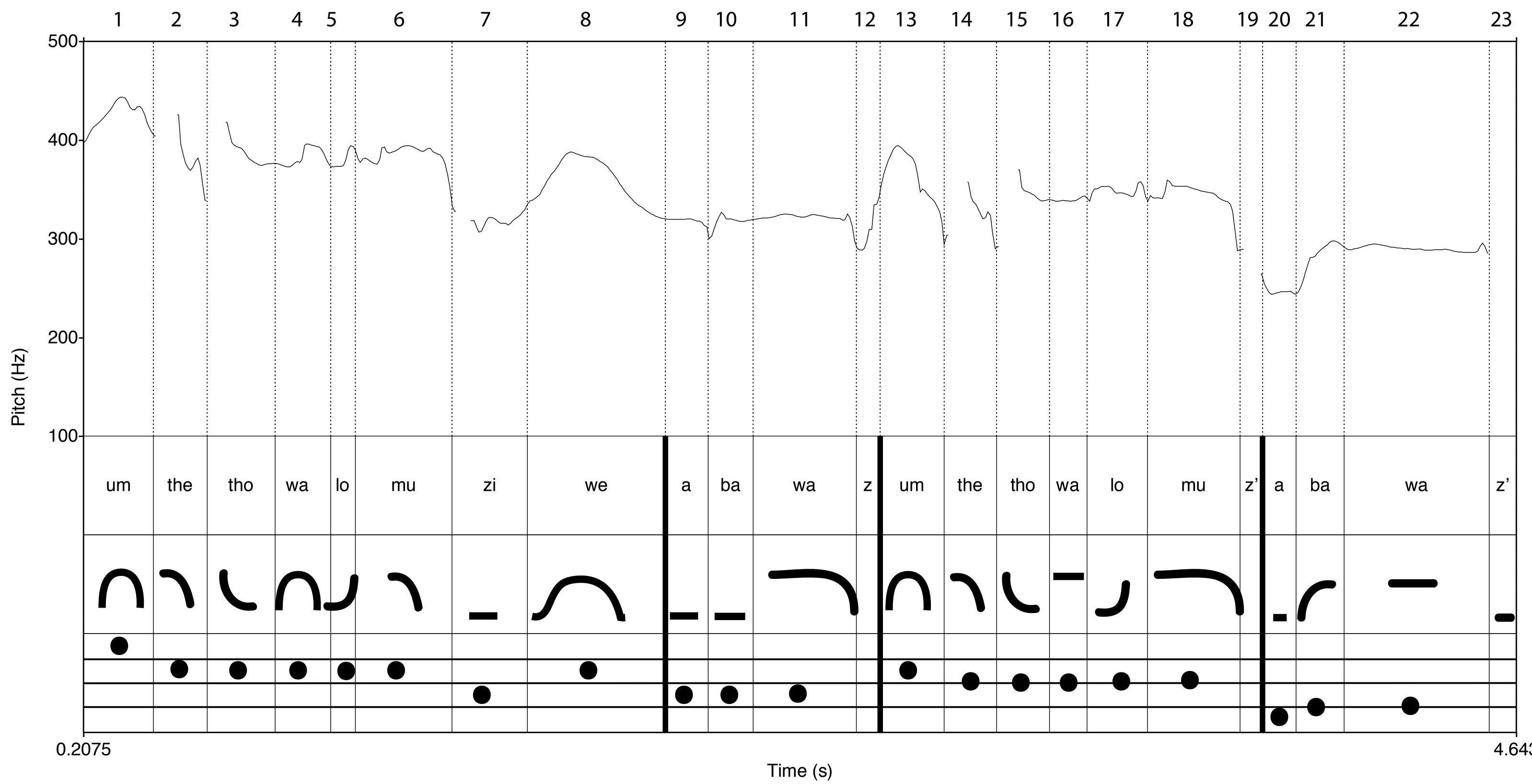
The pitch categories in Figure 1 consist of glides and level tones. Down-glides and up-glides may be concatenated into convex and concave contours. Tones are generally analyzed in terms of High/Low contrasts, although a case may also be made for the inclusion of mid-tones in some descriptions.

These symbols are designed to align with the tonal features of isiZulu. Level tones are used to indicate the contrastive features of High and Low as well as downstep and sometimes mid-tones. Glides indicate the action of depressor consonants. Both glides and contours are used for pitch targets that are realized by movements rather than levels. There is significant variation in the size and gradient of these various glides and contours but the underlying categories are comparable. There are instances where combinations of these categories are concatenated over a single syllable. This is in fact a feature of isiZulu tonology that requires further study because it is often difficult to separate one or more segments due to the gestural or 'gestalt' quality of the articulation. All of these pitch categories are abstract and their combination conveys both linguistic and musical meaning. What has often been overlooked in transcription and analysis of these prosodic features is the extent to which microprosodic fluctuations are perceptual cues that identify points of articulation.

I recognize that pitch reductions of this sort carry with them all the restrictions embodied in their technological rendering. For instance, measurements of fundamental frequency do not correspond precisely to the actual pitch produced by the vocal organs and perceived by the auditory apparatus; it is smoothed out. We must also factor the limitations of digital sampling. I checked on the accuracy of the F_0 tracks and corrected sampling errors and octave jumps. But it must be recognized that pitch tracks facilitate *new* modes of hearing that are artificial. The graphic realization of these sounds mediates and structures our experience. These analytical procedures are heuristic. To hold too closely to a phonetic 'reality' would be to mistake surface structures for cognitive reality. There are also important physiological factors that must be taken into account in this analysis. The rise and fall of subglottal pressure causes acute perturbations in the pitch contour that are independent of supralaryngeal articulatory mechanisms. For instance, the gradual lowering of pitch through a phrase and the concatenation of the range toward the end of a phrase may in part be attribted to physical limitations in the production of sounds. As listeners we compensate for this so that reduced pitch spaces at the ends of phrases are heard according to similar contours and pitch categories as those at beginnings. In other words, we are able to accommodate and scale changes in pitch structure according to our learned categories.

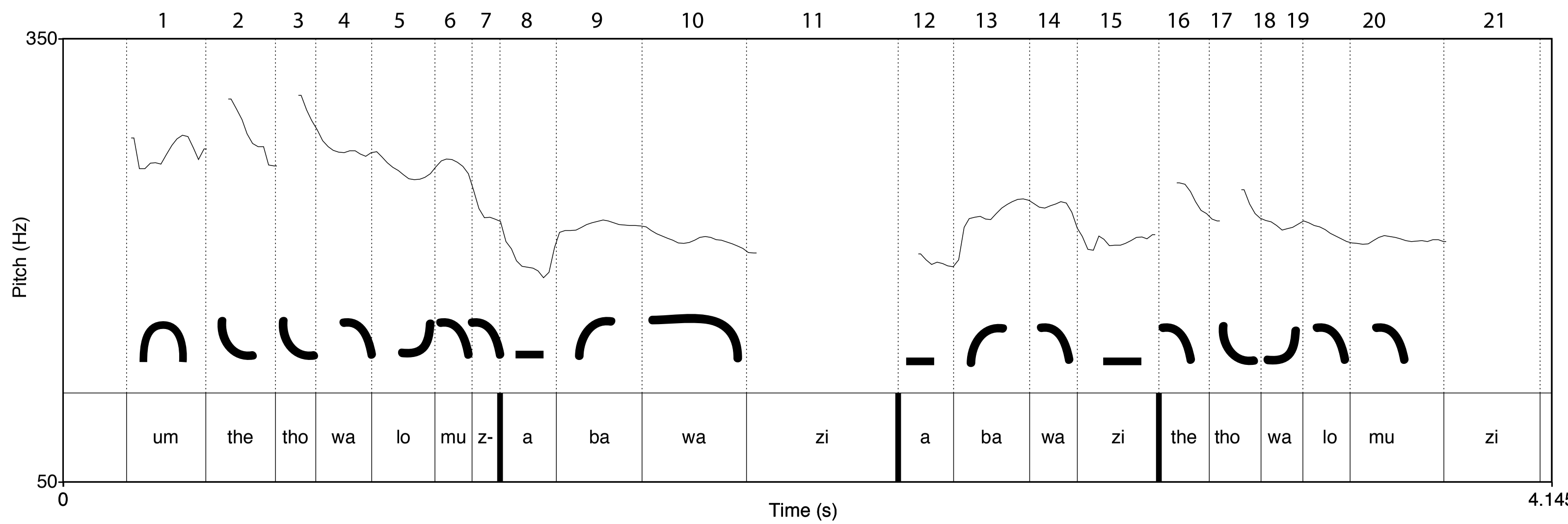
This system of analysis makes several assumptions. First is that we may derive pitch categories from a continuous melodic contour. These categories are combinatorial and are integral to prosodic structure but they are not independent of intonational features, nor of the pitch targets specified in the lower rung of the song graphs. These categoreis are not specified by pitch height but by *movement* (see Chapter 3, Part 2). The use of pitch targets rather than discrete pitches means that a degree of flexibility is allowed for discrepancies within and across the pitch space. This is useful because of the narrowing of pitch space due to depressor features, including the action of consonants, downdrift, and closure.

Example 7.1: *Umthetho* sung by Benzani Dladla



This analysis of *Umthetho* includes four systems: the upper system shows the continuously descending pitch contour (tracking the fundamental frequency). Below is a transcription of the isiZulu lyric divided into segments numbered on the x axis at the top of the graph. The third system is an analysis of the pitch categories corresponding to the individual word segments. These are aligned with the pitch targets in the fourth system at the bottom of the graph.

Example 7.2: *Umthetho* spoken by Benzani Dladla



The spoken pitch contour, pitch categories, and word segments are included in this analysis. There are two main phrases that are separated by the unvoiced 'zi' in segment 11. An important speech tone factor depicted in this graph is the action of depressor consonants. The stop consonants in segments 2, 3, 16, and 17 all result in a pitch lowering effect.

Lyric: *Abawazi umthetho walomuzi.*

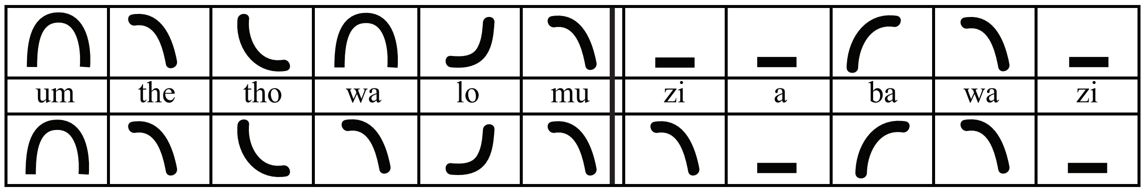
They don't know the law of this homestead.

[Neighbors witness (and pass judgment) on something that's happening in a homestead but don't understand the laws (customs).][256]

There are six pitch targets in this song that extend over an octave range (a **pentatonic** system). Several features of Zulu tonal grammar are shown in this analysis: There is 'downdrift' over the course of each phrase and the cycle as a whole; there are two descending phrases each of which ends on the voiced depressor consonant 'z.' This unvoiced depressor is used consistenly to signal closure; phrase-final pitch categories are longer in duration and lower in pitch than most others in the melody. From a melodic perspective there appears to be no clear tonal center and this is a characteristic of cyclic, interlocking forms. The two phrases could, however, be interpreted as sequential, with the second occupying a slightly reduced pitch range in comparison to the former. Looking at the pitch tracks of examples 7.1 and 7.2 we see considerable similarities. This may be attributed to the prosodic features described here and this is shown in the comparison of pitch categories rendered in Example 7.3.

[256] My thanks to Ndididi Dladla and Landiwe Dladla for assisting with the translation and interpretation of these song texts.

Example 7.3: *Umthetho*, comparison of pitch categories for song and speech[257]



A comparison of the pitch categories in *Umthetho* with song categories at the top and speech categories at the bottom. This illustrates the importance of linguistic factors in conditioning pitch movement. For instance, the action of depressor consonants such as the stop consonant ‘t.’ Notice also that the nasal ‘m’ and fricative ‘z’ have tone-lowering effects. There are two differences in pitch movement that we notice in this comparison. The first is on ‘wa.’ In the sung rendition there is scope for deviation given that no inherent speech tone determinant is in operation. The second is on ‘zi.’ Sometimes this syllable is voiced, in which case there is a depressor effect, but where it is unvoiced we find no such movement.

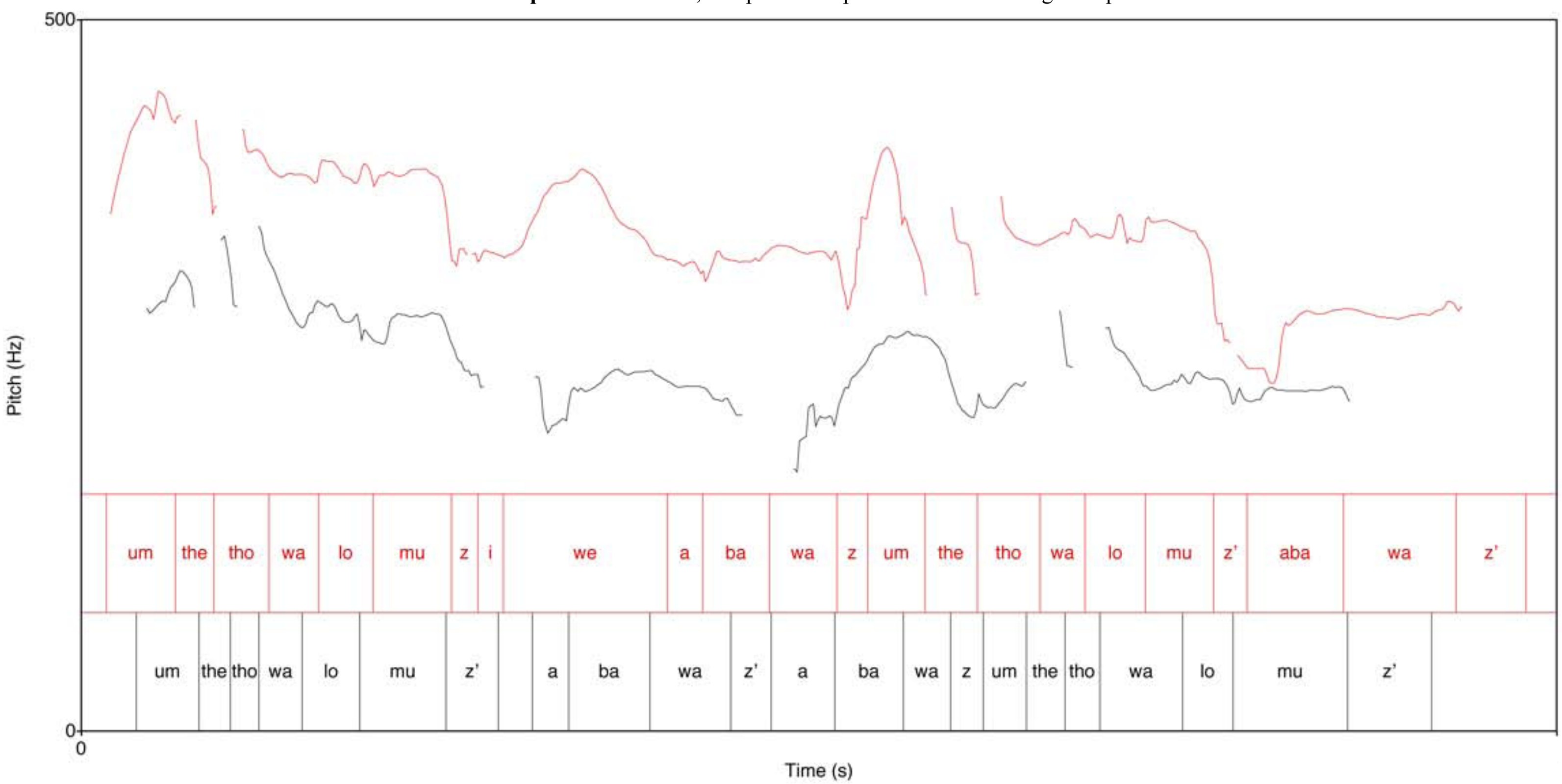
There is a very close correlation between the pitch categories of song and speech. There are several differences of interpretation but the general movement of song and speech contours is consistent in both. This suggests that the pitch structure is closely related to speech tone requirements even if it is not fully determined by them. These requirements include the action of several depressor consonants found in this example, as follows:

- Stops: The syllables ‘the’ and ‘tho’ result in down-glides.
- Fricatives: The depressor function of ‘z’ is clear at both mid and end points in this song. This is true even where it is unvoiced.
- Nasals: ‘m’ takes a depressor role in the segments ‘um’ and ‘mu.’ In ‘um’ it functions as a depressor on ‘u’ whereas in ‘mu’ it functions as a depressor on the preceding pitch contour ‘lo.’

Another method by which we may compare the song and speech contours of *Umthetho* is to align their overall fundamental frequency contours, as in Example 7.4.

[257] Note that these segments have been chosen on the basis of their comparability. The double bar lines show divide different parts of the utterance that are not continuous.

Example 7.4: *Umthetho*, comparison of pitch contours for song and speech



The pitch track of *Umthetho* sung is in red and the pitch track of *Umthetho* spoken is in black. This analysis shows the broad similarities in the overall pitch contours of the sung and spoken versions, and it reinforces the analysis of pitch categories depicted in example 7.3. For the purposes of comparison, the pitch tracks have been formatted to align even though this distorts the time domain on the X-axis.

Song/speech pitch parallelism is evident not only in the pitch categories but also in the intonation or overall melodic contour.[258] There is a significant difference in overall pitch height in these two renderings. The spoken pitch contour (black) is far lower than the sung contour (red). Even so, other similarities include the general pitch movement when we compare syllable by syllable. The overall contour runs in parallel until the two lyrics shift in the middle section with words reversed.

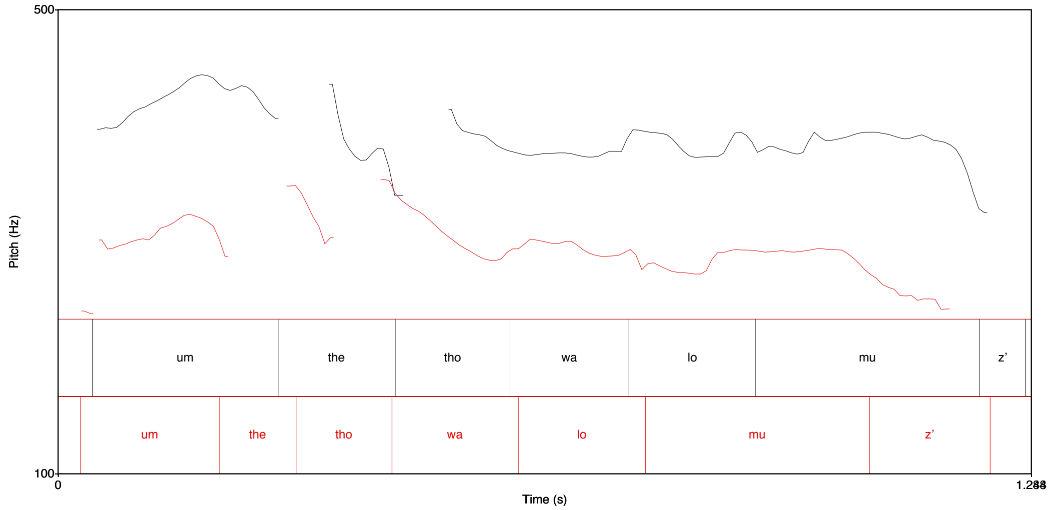
The next three figures (4, 5 and 6) show three separate performances of the first phrase of *Umthetho* by three different singers. These also indicate the close alignment of speech and song melody. It is highly significant that the contours of these three performances show similarities not so much in absolute pitch as in general direction of pitch movement; that is, in their gradience. Throughout we find that the shapes of the song melody are closely aligned to the shapes of the intonation contour. This parallelism is also apparent in the downdrift or declination of the melodic contour from an early highpoint to a low final tone.

Comparative Analysis

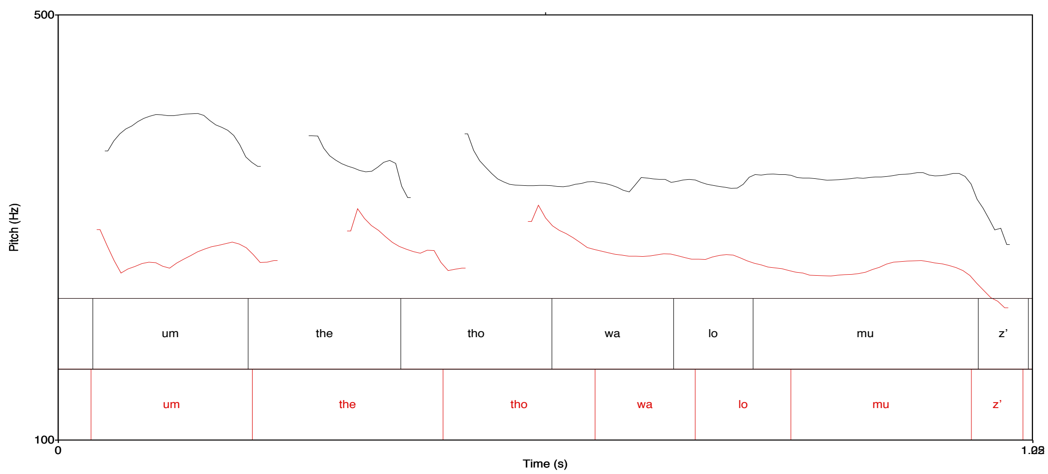
Figures 4, 5, and 6 are pitch tracks of performances by three different girls of the same phrase sung (black line) and spoken (red line). There is considerable variation in pitch height in performances of the same song, an observation that applies in the following three ways:

[258] Note that the time dimension has been marginally distorted in order to fit both contours onto a single system for comparative purposes.

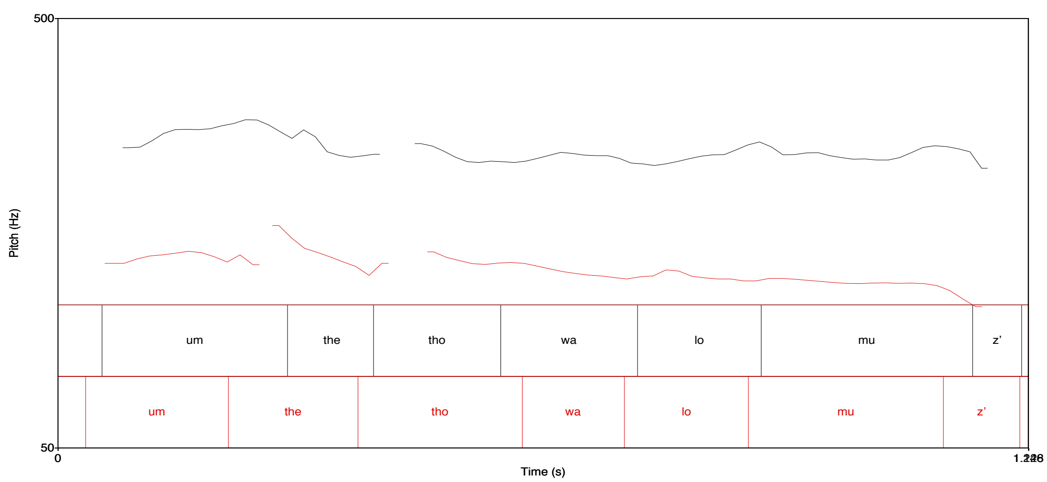
Example 7.5: *Umthetho*, first phrase sung (black) and spoken (red) by Benzani Dladla.



Example 7.6: Umthetho, first phrase sung (black) and spoken (red) by Landiwe Dladla.



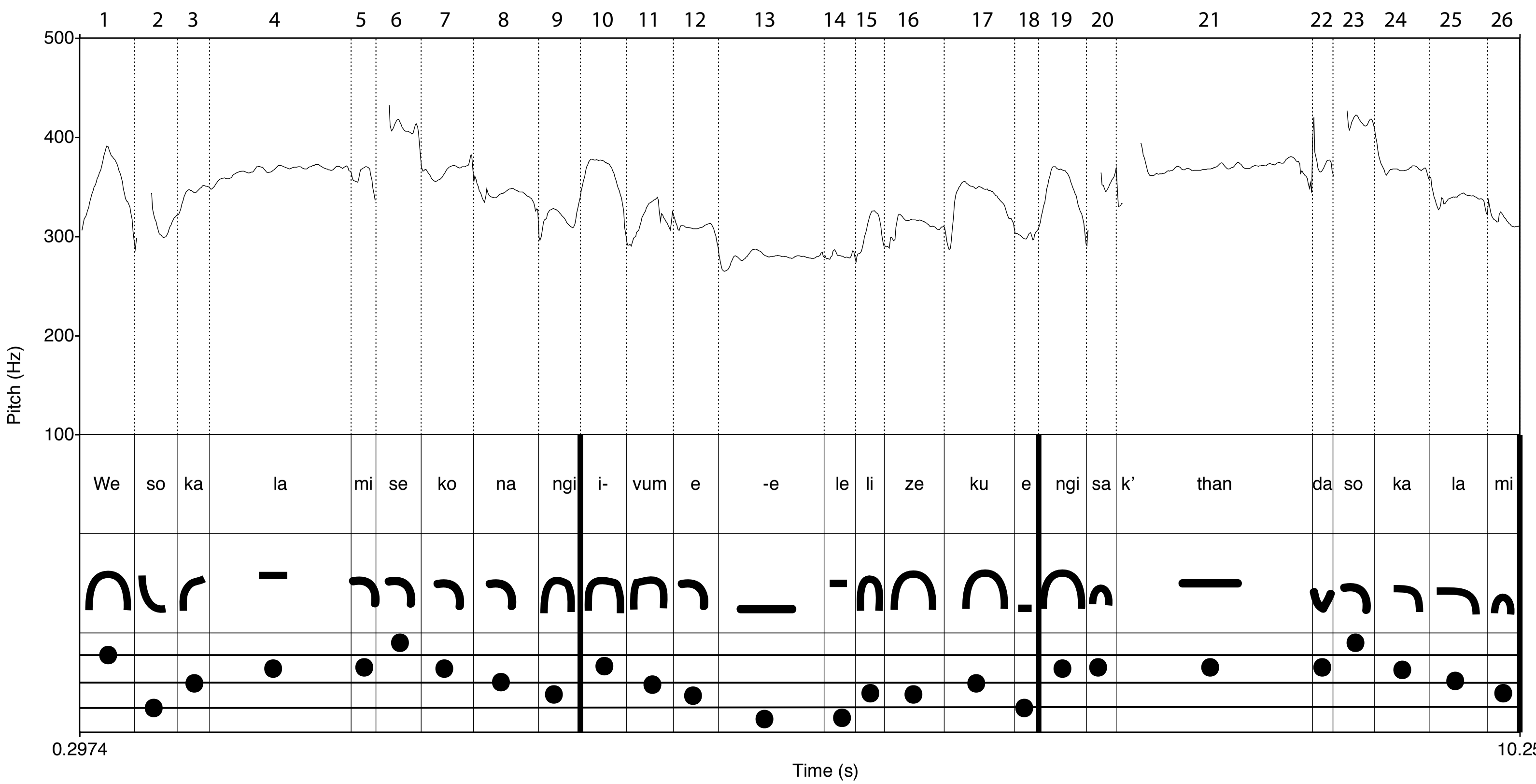
Example 7.7: Umthetho, first phrase sung (black) and spoken (red) by Kwenzekile Dladla.



Each of the singers begins both the spoken and the sung phrases at a different pitch height but there is wide variation in the pitch contours and targets. It is only the gradient or overall pitch profile that is consistent across the three performances. It is important to note that even though the phrase is sung at different pitch heights it remains recognizable based on this gradient alone. This is strong evidence for a theory of pitch contour that supplants a system of discrete pitches or even pitch targets with a more flexible cognitive system. It also supports the claim, outlined in Chapter 1, that a crucial feature of the human cognitive endowment for music is transposability.

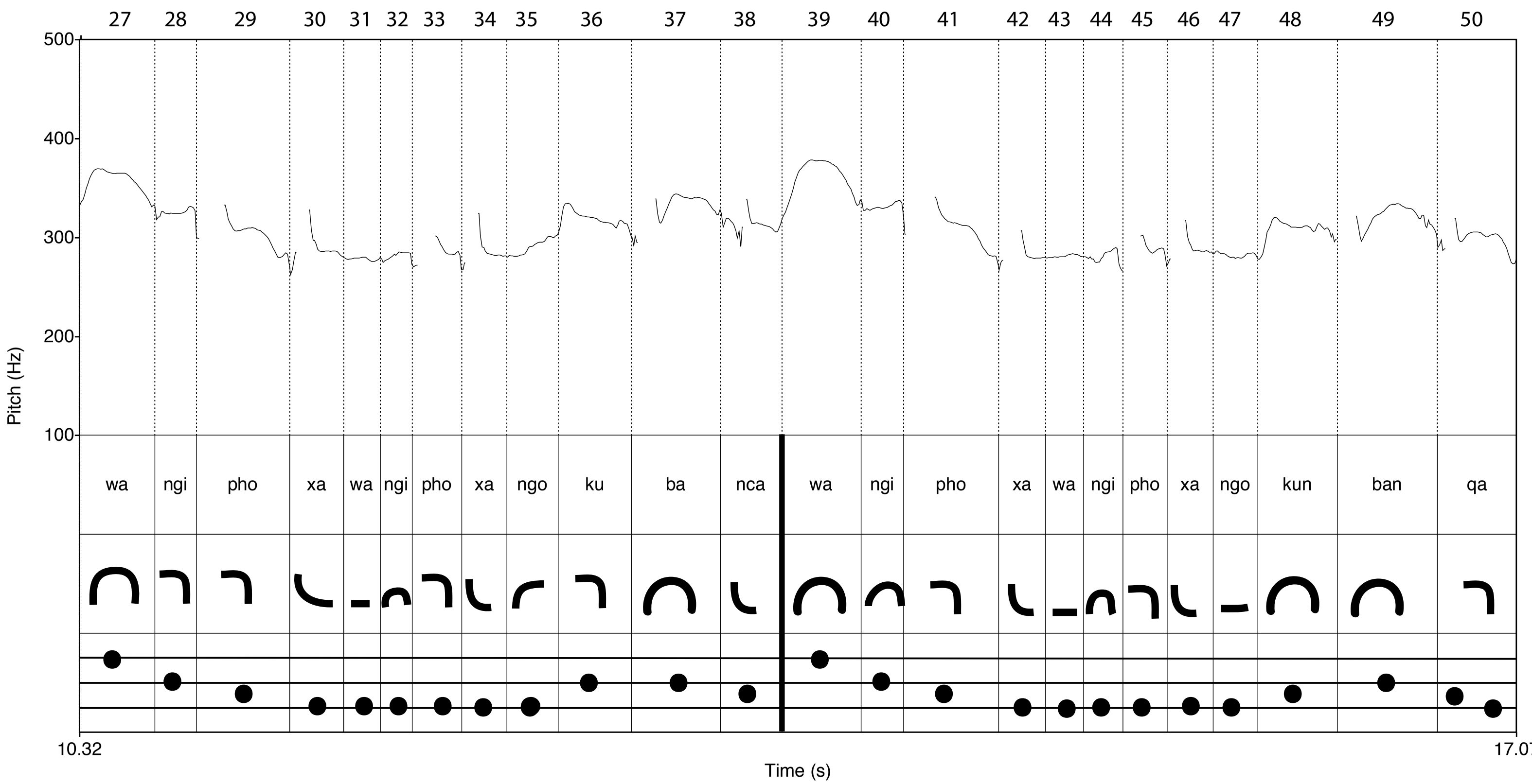
These several analyses of *Umthetho* detail many important aspects of pitch structure in Zulu song prosody. Less detailed is provided on the remaining nine songs but we may observe similar pitch categories and relationships in operation in each case. Rather than describe each of these elements exhaustively I have noted only those points that are unusual or which demonstrate an important process. A discussion and summary of findings follows.

Example 8.1: *Wesoka Lami*, sung by Benzani Dladla (Part 1)

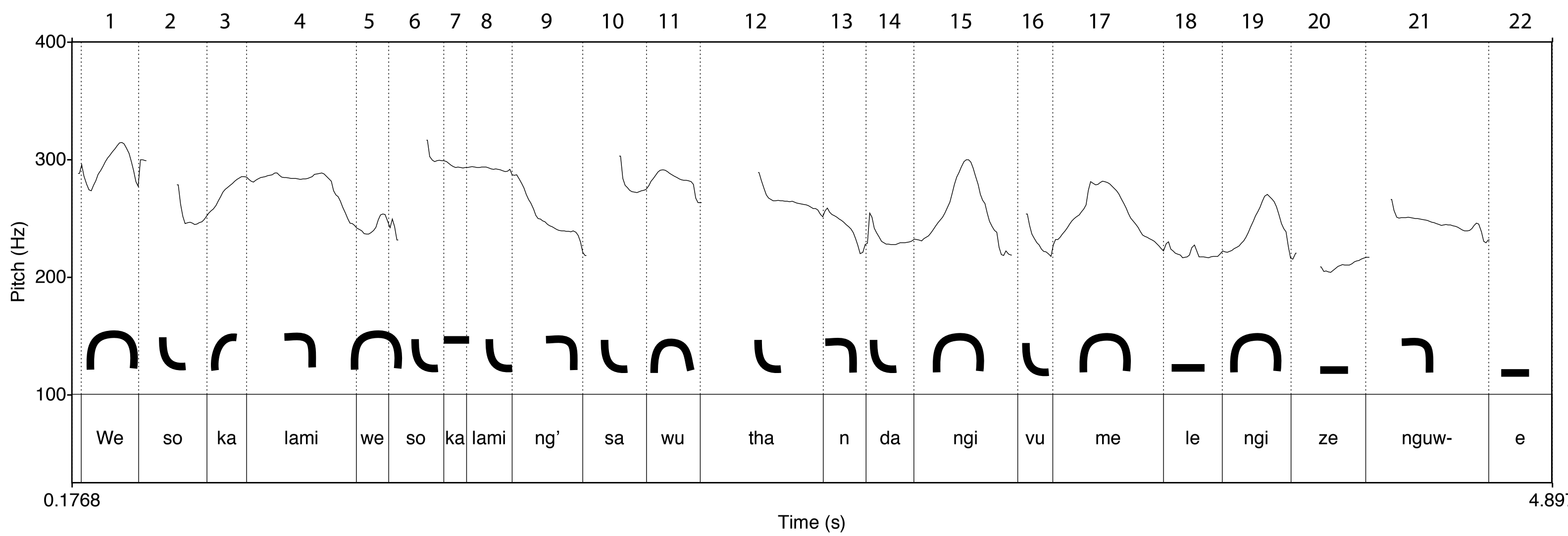


This sung performance of *Wesoka Lami* is divided into the same four systems as Example 7.1 above. The song consists of seven pitch targets spread over a range of less than an octave (i.e., a **heptatonic** system). There are four phrases with rising/falling contours that comprise a cycle. In this song we see how the singer has internalized the call-and-response pattern and interrupts the 'response' to return to the 'call' at segments 22-23.

Example 8.1: *Wesoka Lami*, sung by Benzani Dladla (Part 2)



Example 8.2: *Wesoka Lami* spoken by Benzani Dladla



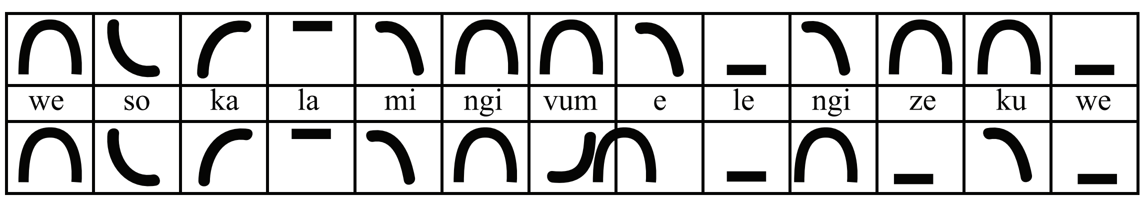
This excerpt shows a gradually constricting pitch-span over the course of the utterance. This is also a feature of the sung version. Due to the short duration of the excerpt, it seems unlikely that this narrowing is due to physical restrictions on production. An alternative explanation is that phrase-level boundaries are marked by declination and pitch span reduction, and that these may not be necessary to the accurate identification of a melodic percept.

Lyric: *Wesoka lami ngisakuthanda. Wangiphoxa ngokungifebela.*

My boyfriend, I still love you. You have disappointed me by having an affair.

This song consists of seven pitch targets spread over a range of less than an octave, with a **heptatonic** system. There are four phrases with rising/falling contours that comprise a cycle. In this song we see how the singer has internalized the call-and-response pattern and interrupts the 'response' to return to the 'call.' This is neatly demonstrated by the comparison of song and speech contours below.

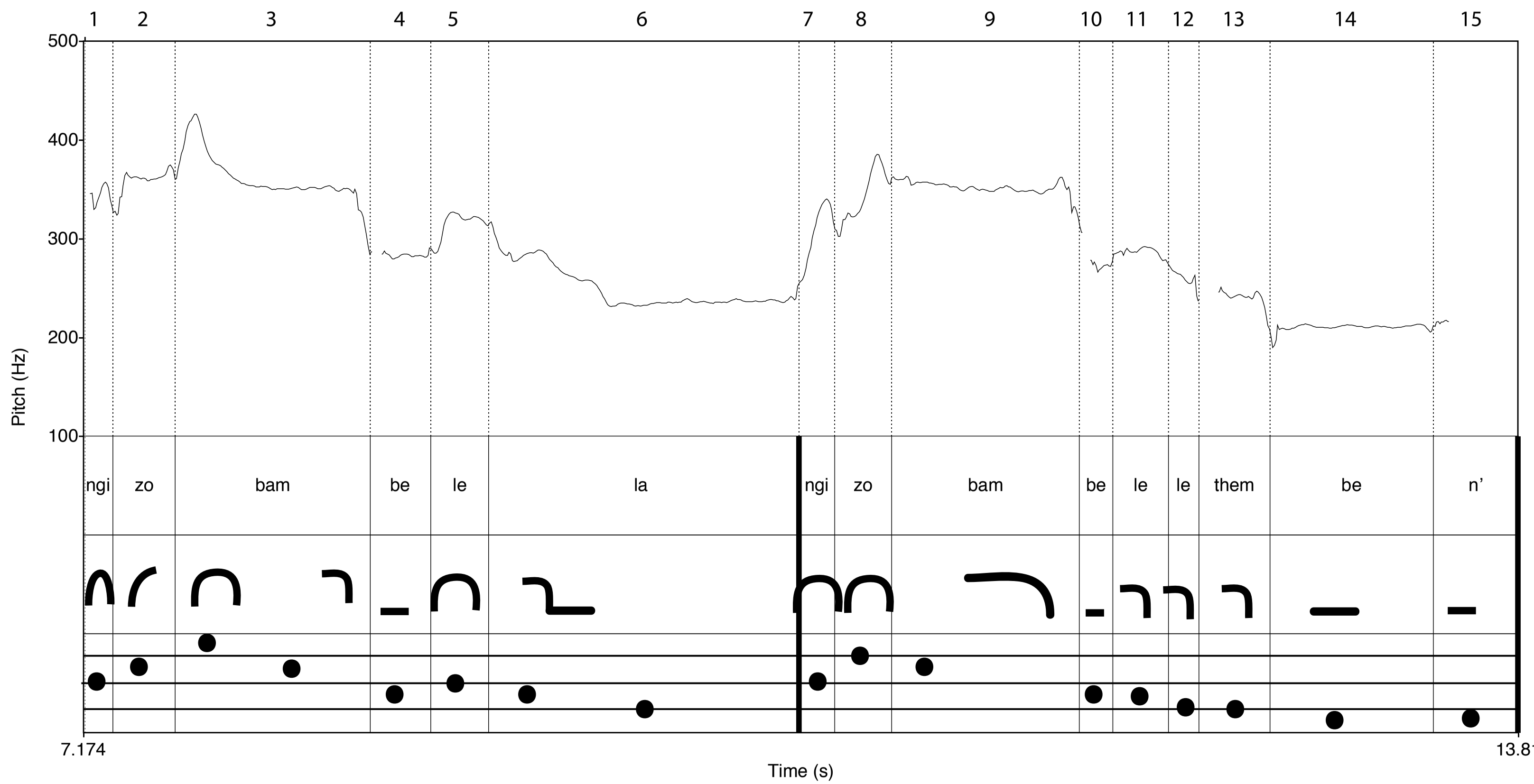
Example 8.3: *Wesoka Lami*, comparison of pitch categories for song and speech



Notice that the pitch categories are very similar, except for the segments 'vum-e' and 'ze-ku.' In the former case there are no strong speech-tone determinants, while in the latter we see how the depressors 'z' and 'k' operate normally. A voiced 'e' is present in the spoken version of the segment 'ze,' and is what accounts for the level pitch category rather than a falling contour.

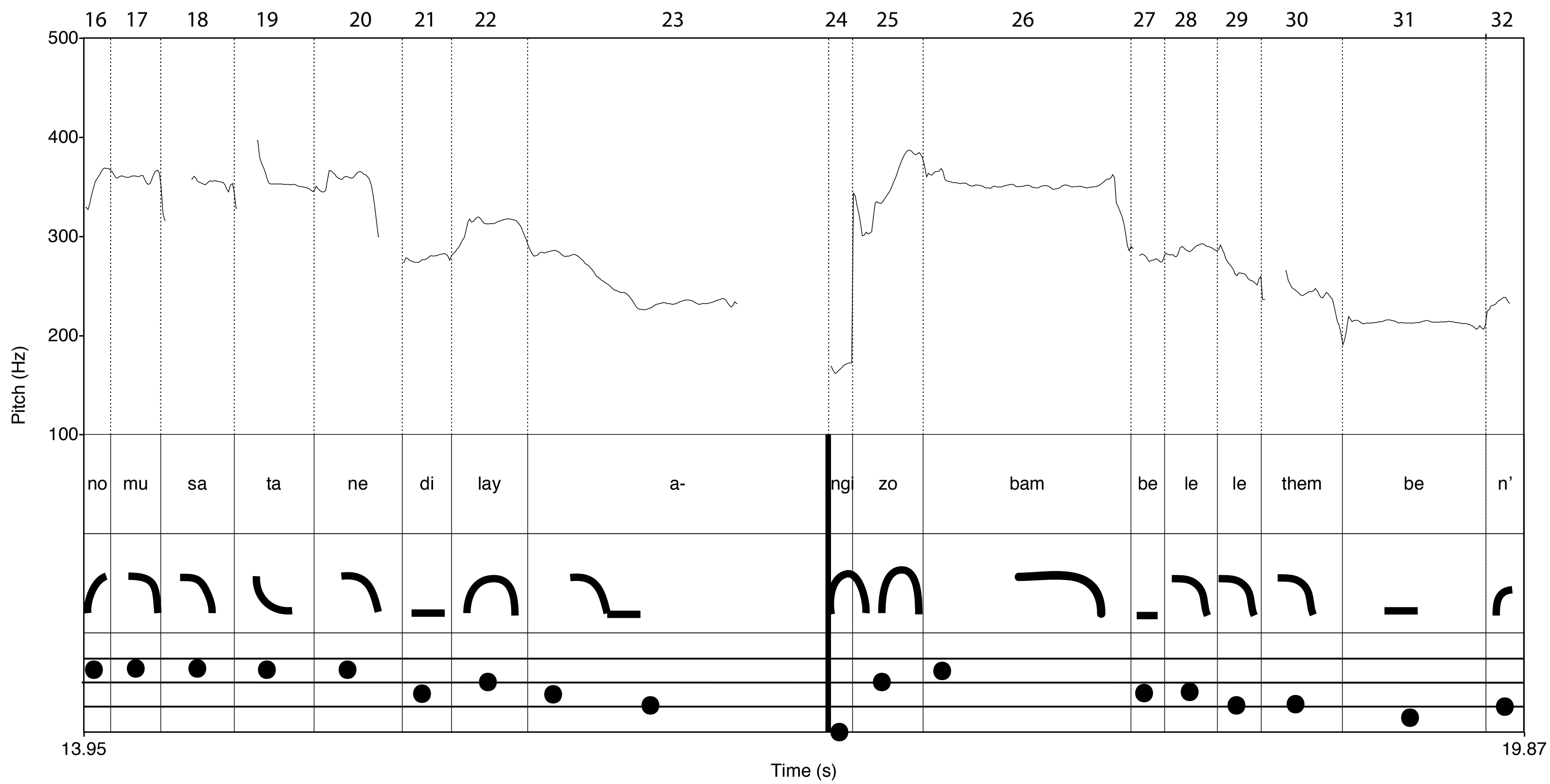
The pitch categories of the song and speech contours are consistent, but not precisely the same in this comparison. The first four segments ('We-so-ka-lami') are realized in a very similar fashion. The numerous convex pitch categories in this song require some explanation because they are not all reducible to speech-tone elements. In some cases the rise is conditioned by a nasal ('ng') but more often the sharp rising-falling pitch contour is realized on a vowel. This suggests that it is a melodic embellishment made for expressive rather than linguistic or lexical-functional purposes.

Example 9.1: *Uzobambelela,* sung by Landiwe Dladla (Part 1)

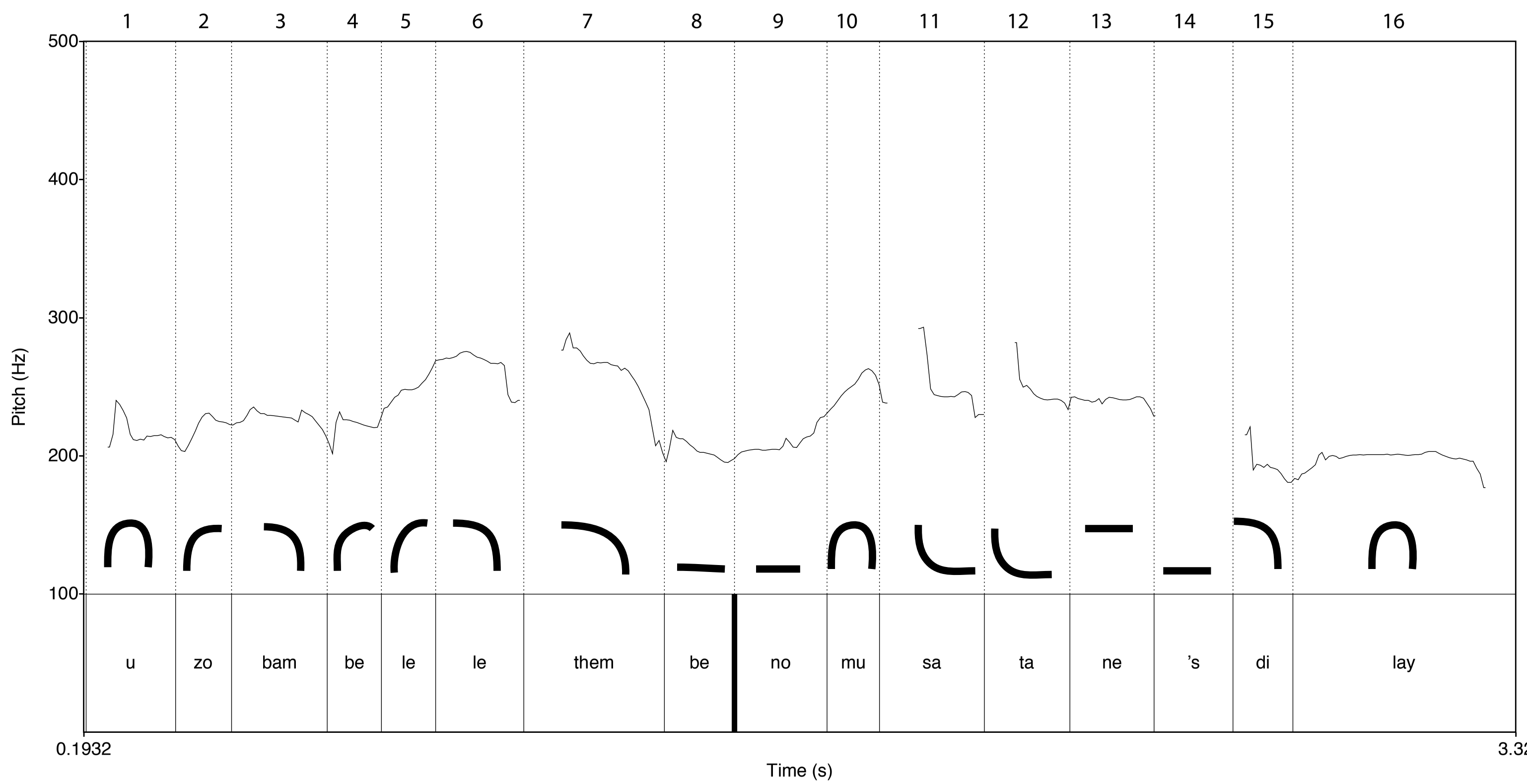


There are eight pitch targets that range over more than an octave (i.e., a **heptatonic** system). The four phrases each have descending contours and are broken by longer durational units at endings. Phrase-final notes are also lowest in pitch even at the end of individual phrases: A (1-6) and B (7-15), C (16-23) and D (24-32). Phrases A, B, and D consist of the same initial syllables and the pitch contours have similar shapes. However, despite consistency of melodic contour and textual content (implying speech tone), the pitch height varies considerably. Notice how the first high tone in segment 3 (phrase A) has the highest frequency followed by a gradual descent over the course of the cycle (as well as a reduction in pitch range, of course).

Example 9.1 Continued: *Uzobambelela*, sung by Landiwe Dladla (Part 2)



Example 9.2: *Uzobambelela* spoken by Landiwe Dladla

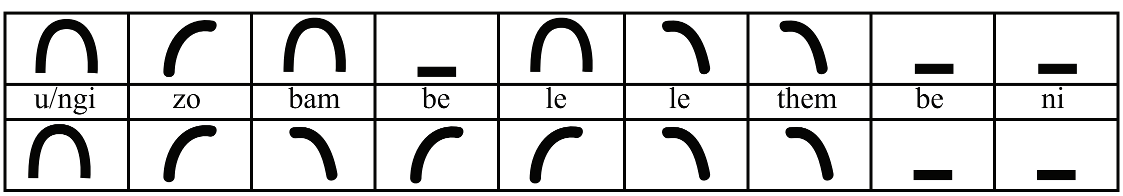


This example shows the depressor action of nasals at segments 3 (‘bam) and 7 (‘them’). It also demonstrates the way in which vowels like ‘u’ (segment 1) are realized in multiple different pitch movements depending on context. They tend to be less consistent in their pitch movements than those determined by speech tone requirements.

Lyric: *Ngizobambelela ethembeni noma usathane engidilaya.*
I will hold onto hope even though the devil surrounds me.

There are eight pitch targets that range over more than an octave, suggesting a **heptatonic** system. The four phrases each have descending contours and are broken by longer durational units at endings. Phrase final notes are also lowest in pitch even at the end of individual phrases: A (1-6) and B (7-15), C (16-23) and D (24-32). Phrases A, B, and D consist of the same initial syllables and the pitch contours have similar shapes. However, despite consistency of melodic contour and textual content (implying speech tone), the pitch height varies considerably. Notice how the first high tone in segment 3 (phrase A) has the highest frequency followed by a gradual descent over the course of the cycle (as well as a reduction in pitch range, of course).

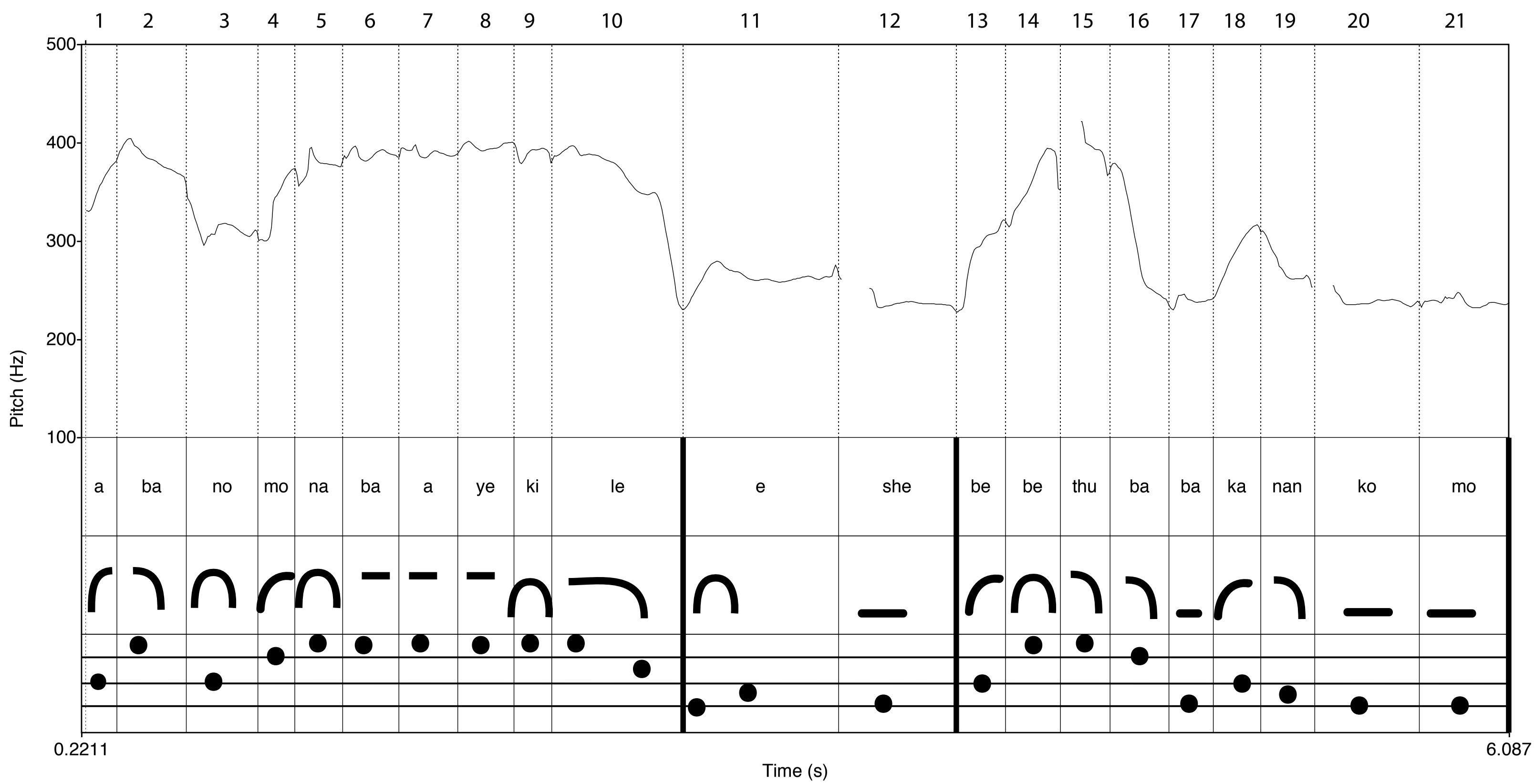
Example 9.3: *Uzobambelela*, comparison of pitch categories for song and speech



The categories of speech and song are nearly identical in this comparison. The discrepancy on 'be' may be explained by the fact that there is a brief onset to the level tone in the spoken contour.

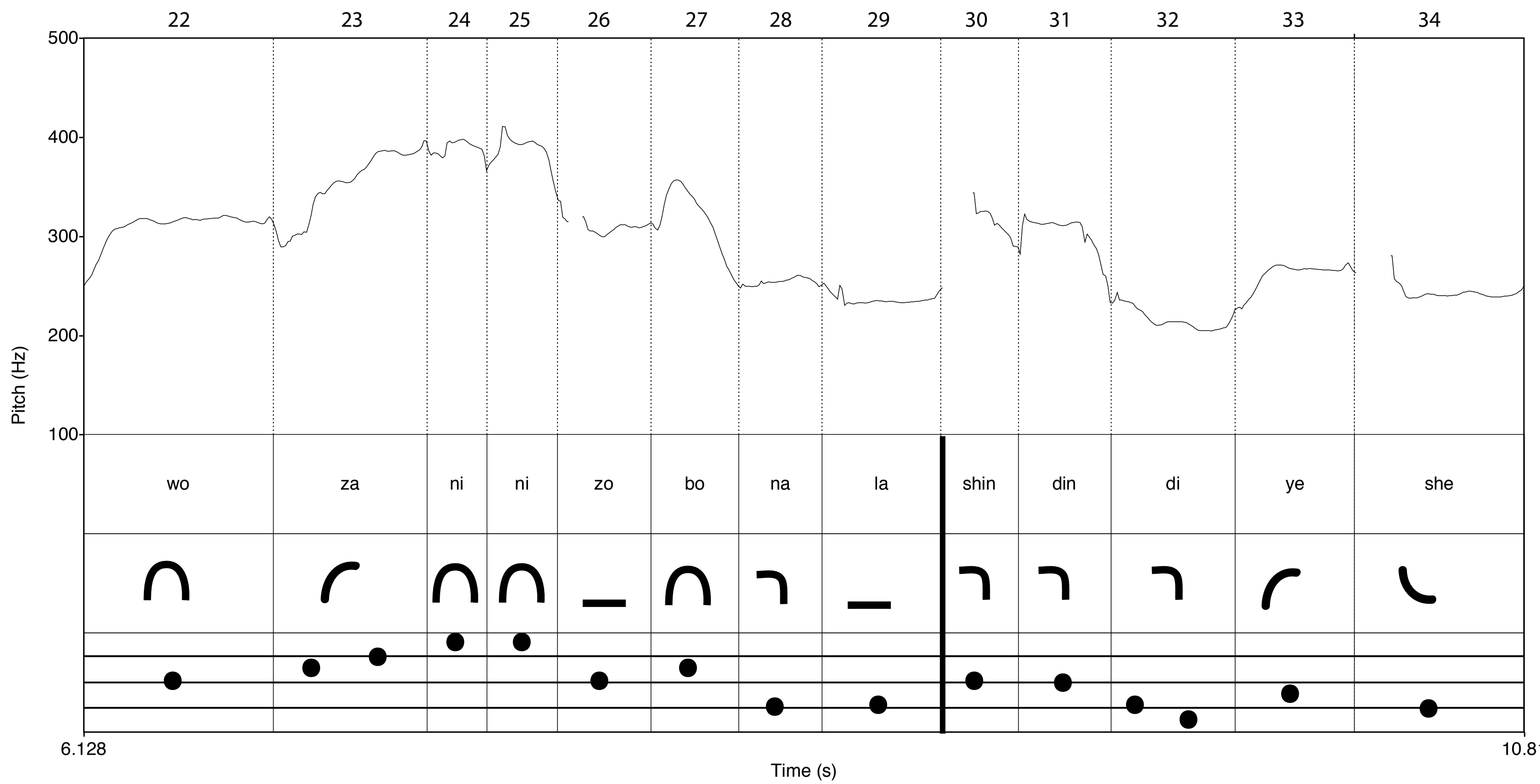
In this comparison of song and speech pitch categories we notice the effects of depressor consonants in several segments. For instance, the nasal 'm' has a clear depressor function on the word-initial syllables in 'bam' and 'them.' The nasal 'n' in segments 15 and 32 is voiced and in both these cases results in a rising approach (up-glide). The stop consonant ‘t’ has a depressor function while for ‘b’ the situation is less consistent. It may be explained by implosive ‘b’ not affecting a depressor function. This example also shows the sort of *portamento* I described in the previous example where the pitch contour moves independently of its linguistic determinants. This usually occurs on vowels or liquids held for longer durations. See, for instance, segments 6 and 23, where significant fluctuation in the pitch contour occurs expressive purposes.

Example 10.1: *Abanomona* sung sung by Benzani Dladla (Part 1)

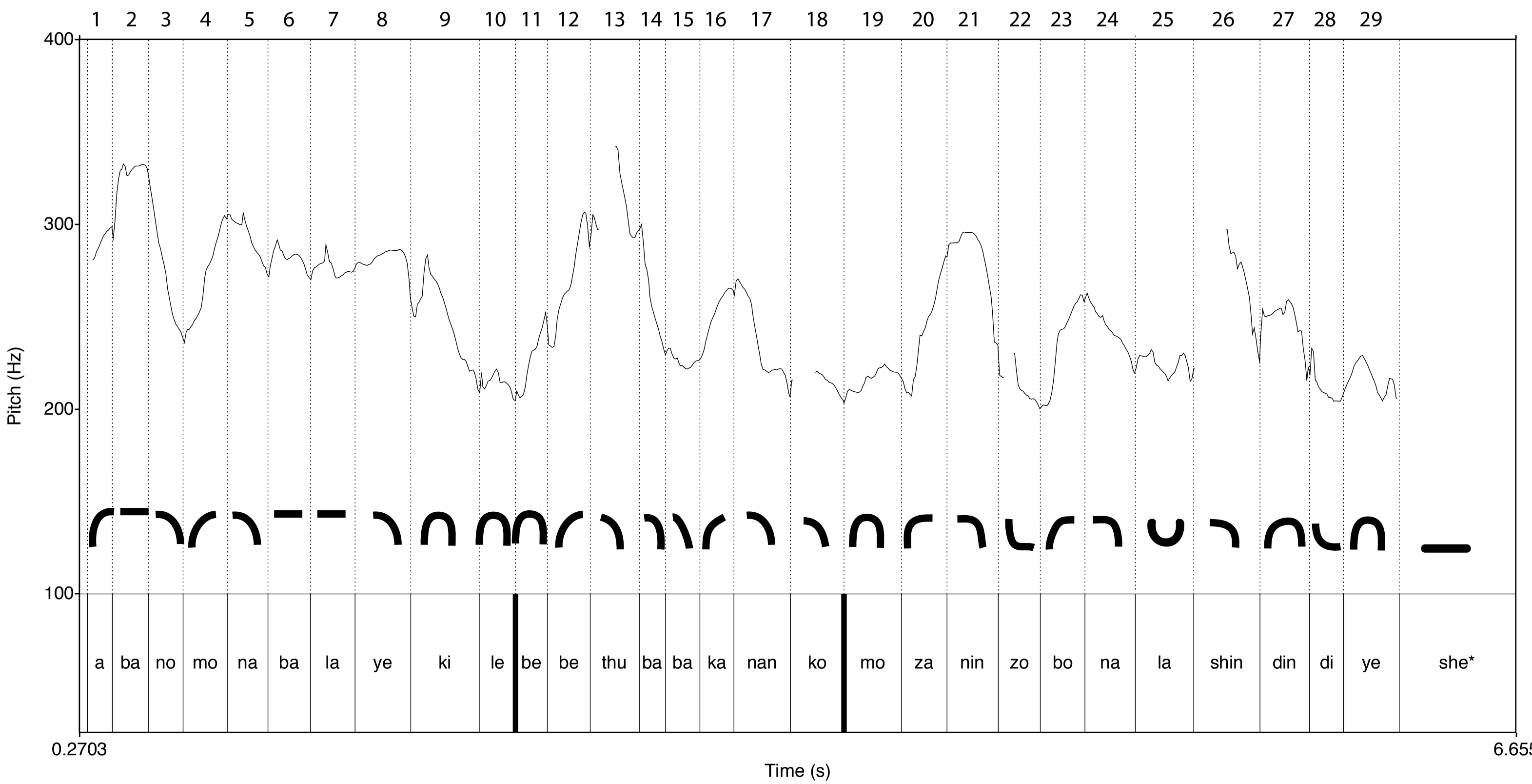


There are seven pitch targets shown in this analysis of Abanomona (i.e. a **hexatonic** system) divided into two phrases in the cycle: A (1-21) and B (22-34). There is distinctive downdrift with high points reached early in the phrase, and then a tapering down to the sustained level pitches at phrase-final positions, signaling closure.

Example 10.1 Continued: *Abanomona* sung by Benzani Dladla (Part 2)



Example 10.2: *Abanomona* spoken by Benzani Dladla



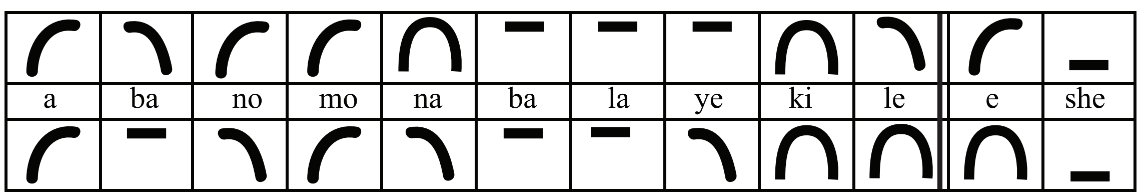
The gradual downdrift and narrowing of pitch span evident in the spoken lyrics to *Abanomona* is consistent with that of the sung version (see 10.1). The depressor action of nasals (5, 17, 21, 24, 26, 27) and stops (e.g., segment 9) (see also 10.3 for comparison of selected song and speech pitch movements) shows that these are phonetic features operational across domains.

Lyric: *Abanomona balayekile yeshe. Bebethi ubaba akanankomo, wozani nizobona shindindiyeshe.*

Those who have jealousy are disappointed. They say my father doesn't have cows, come and you will see.

There are seven pitch targets in this song that are spread over an octave range. This is a **hexatonic** system with two distinct phrases in the cycle: A (1-21) and B (22-34). There is distinctive downdrift with high points reached early in the phrase and then a tapering down to the sustained level pitches at phrase-final positions, signaling closure.

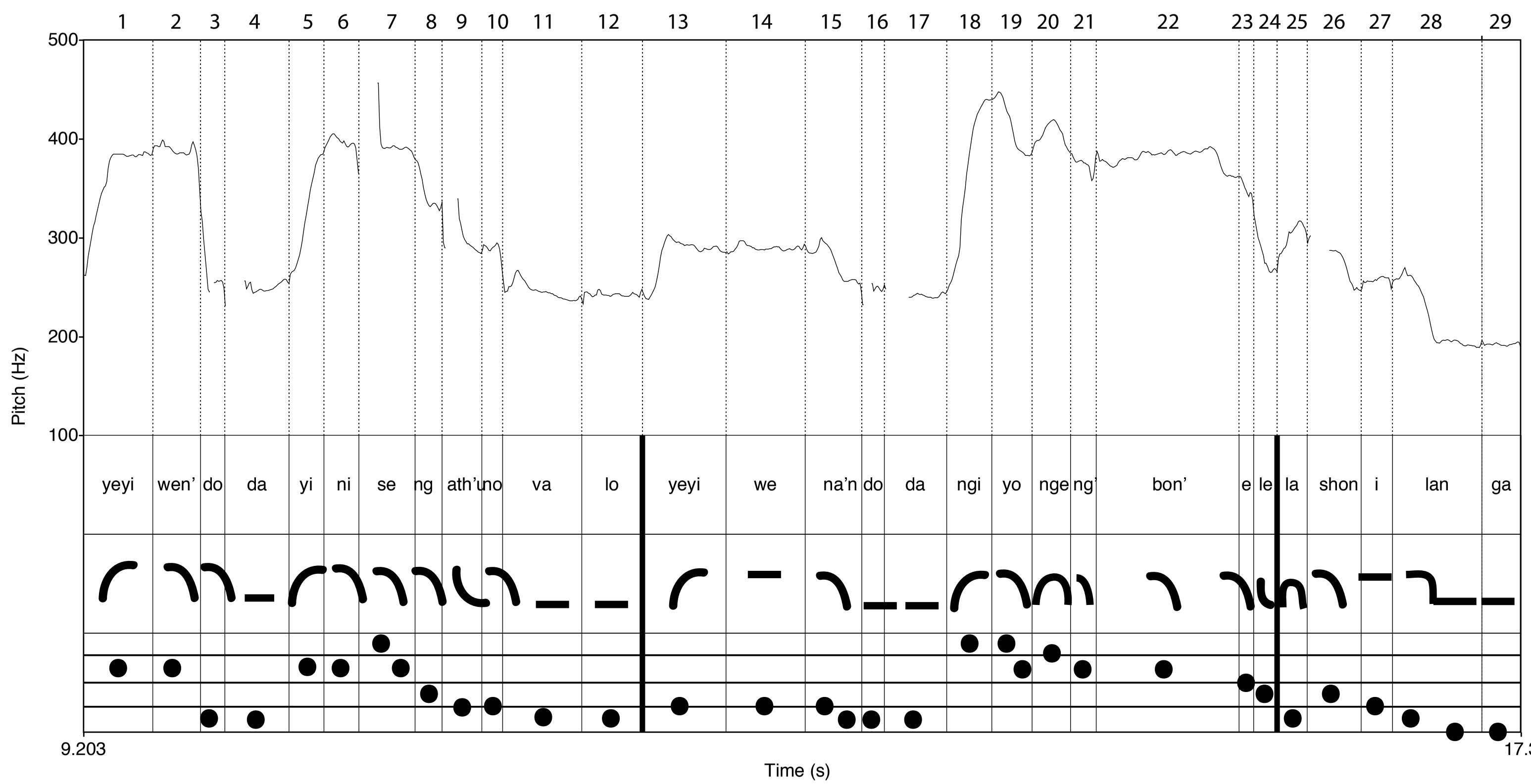
Example 10.3: *Abanomona*, comparison of pitch categories for song and speech



This comparison shows, once again, how similar the pitch categories for song and speech are. One important deviation occurs on the second segment where 'ba' is a down-glide in the sung version, but an up-glide in the spoken version.

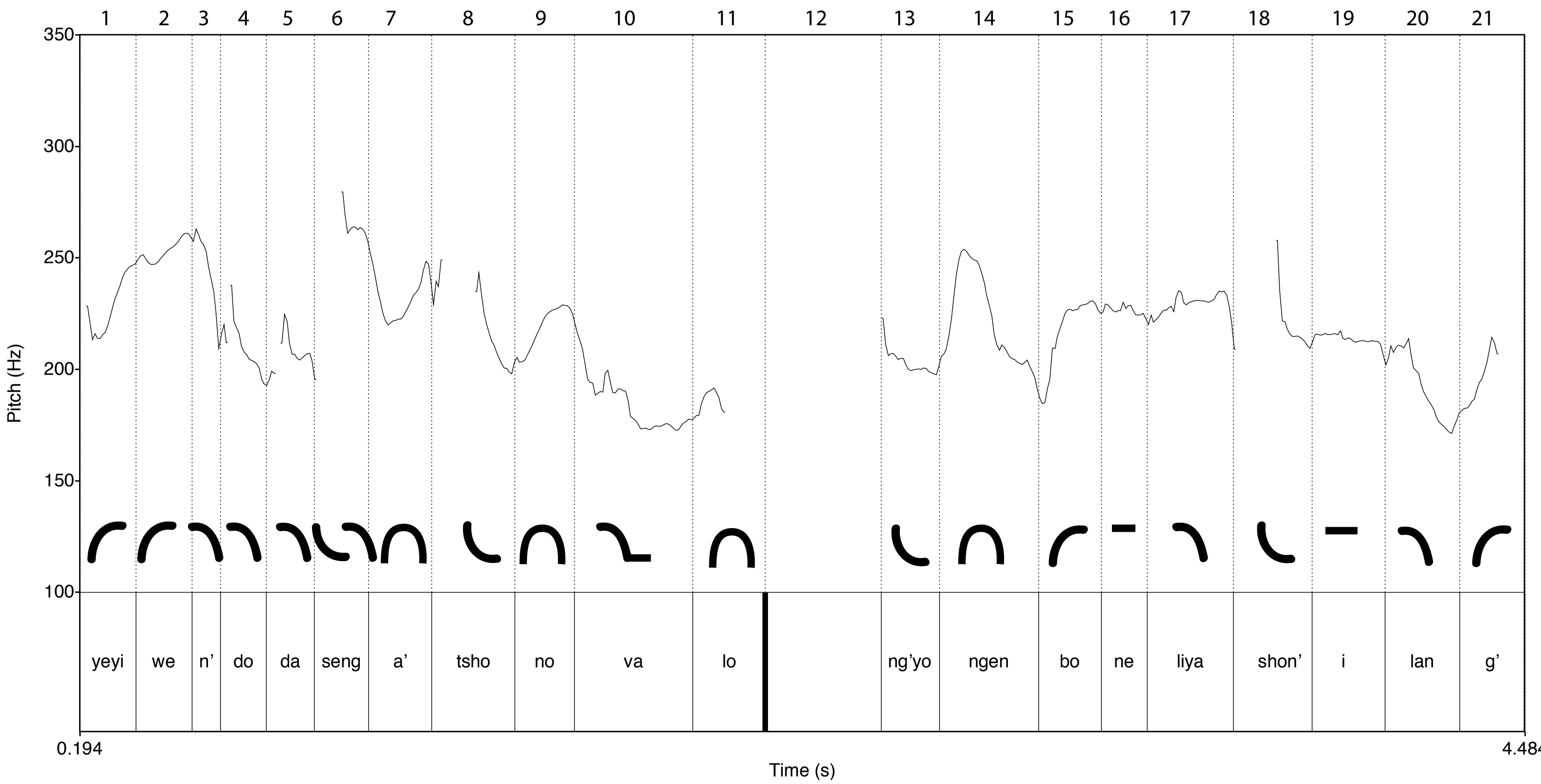
The pitch contours for song and speech targets are very similar. This is interesting because there are so many contour tones. Since contours and glides are *not* commonly used for expressive purposes in speech (certainly not to signal stress; see Ladd 2008: Chapter 1), we must assume that these categories are determined instead by speech-tone requirements. They are not, after all, the product of intonational features either since these display gradience within categories rather than across them.

Example 11.1: *Uvalo* sung by Landiwe Dladla



There are eight pitch targets in this song and they are spread over a range of more than an octave (i.e., a **heptatonic** system). Two main phrases A (1-12) and B (13-24) divide the cycle. A comprises a single melodic idea whereas B is divided into three subphrases (13-17; 18-24; 25-29). There is evidence of 'downdrift' once again, as well as a narrowing of the pitch-span. Both phrases descend from high- to low points, and the phrase ends at the lowest level pitch in the cycle.

Example 11.2: *Uvalo* spoken by Landiwe Dladla



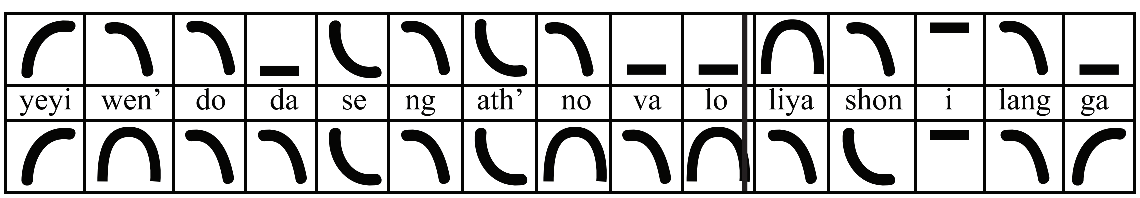
Sibilant onsets such as 's' and 'sh' (segments 6, 8, and 18) have a characteristic onset. I have used a convex down-glide as a descriptor in each. The interpretation is phonetic because sibilant are characterized by high frequencies. The descent to the vowel is therefore precipitous and results in the down-glide pitch movement found in this example, and many others in discussed in this chapter.

Lyric: *Yeyi wena indoda, yini angathi unovalo. Ngiyoke ngibone liyashona ilanga.*

Hey you man, what are you scared of? I will wait and see because the sun is setting.

There are eight pitch targets in this song and they are spread over a range of more than an octave making this a **heptatonic** system. Two main phrases A (1-12) and B (13-24) divide the cycle, followed by a short coda. A comprises a single melodic idea whereas B is divided into two subphrases (13-17; 18-24). There is evidence of 'downdrift,' once again, as well as a narrowing of the pitch-span. Both phrases descend from high- to low-points and the phrase ends at the lowest level pitch in the cycle.

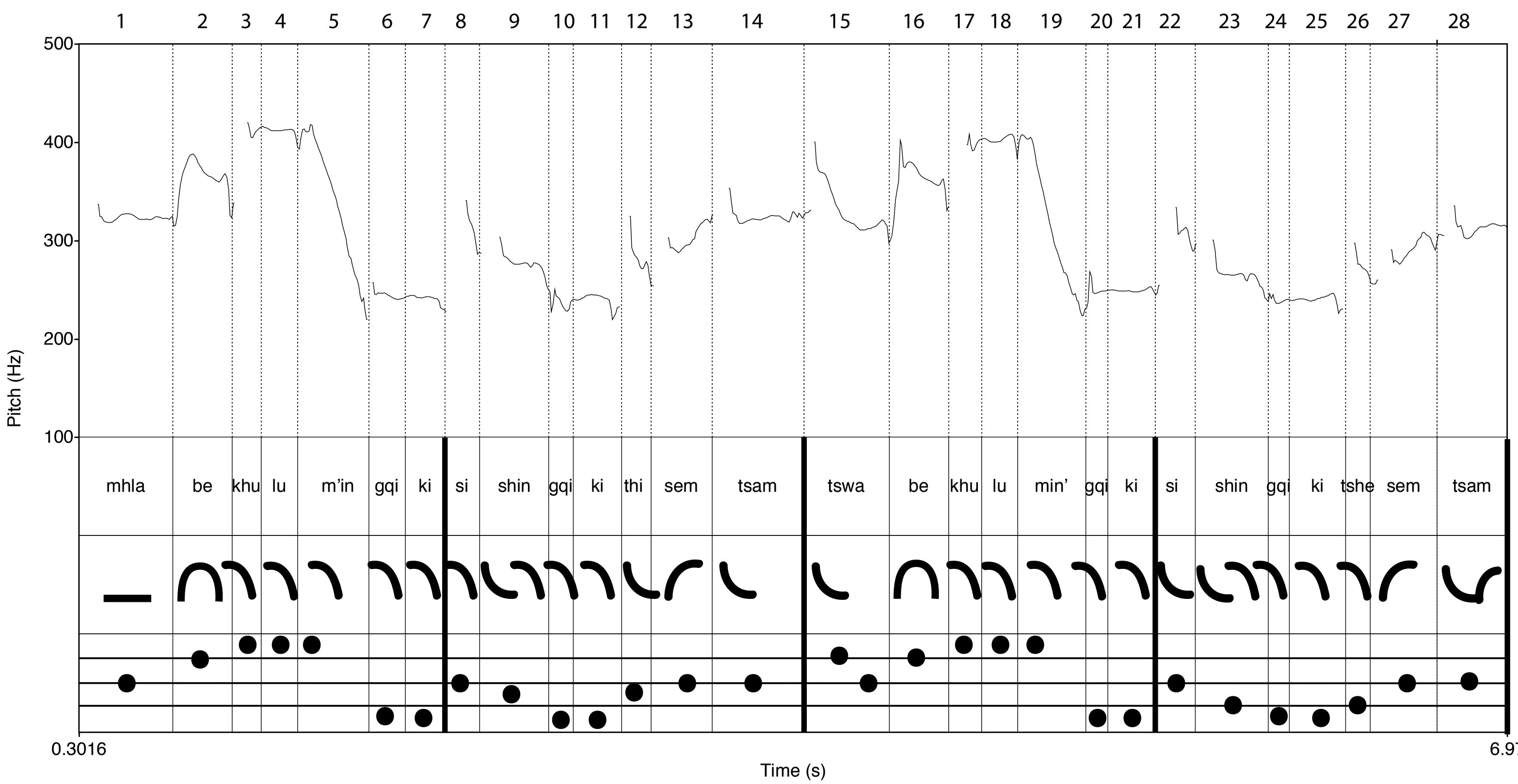
Example 11.3: *Uvalo*, comparison of pitch categories for song and speech



Pitch categories for song and speech are very similar except for a few additional glides (e.g. 'wen' in segment 2. This also explains differences regarding level pitch categories as opposed to glides (e.g. 'da').

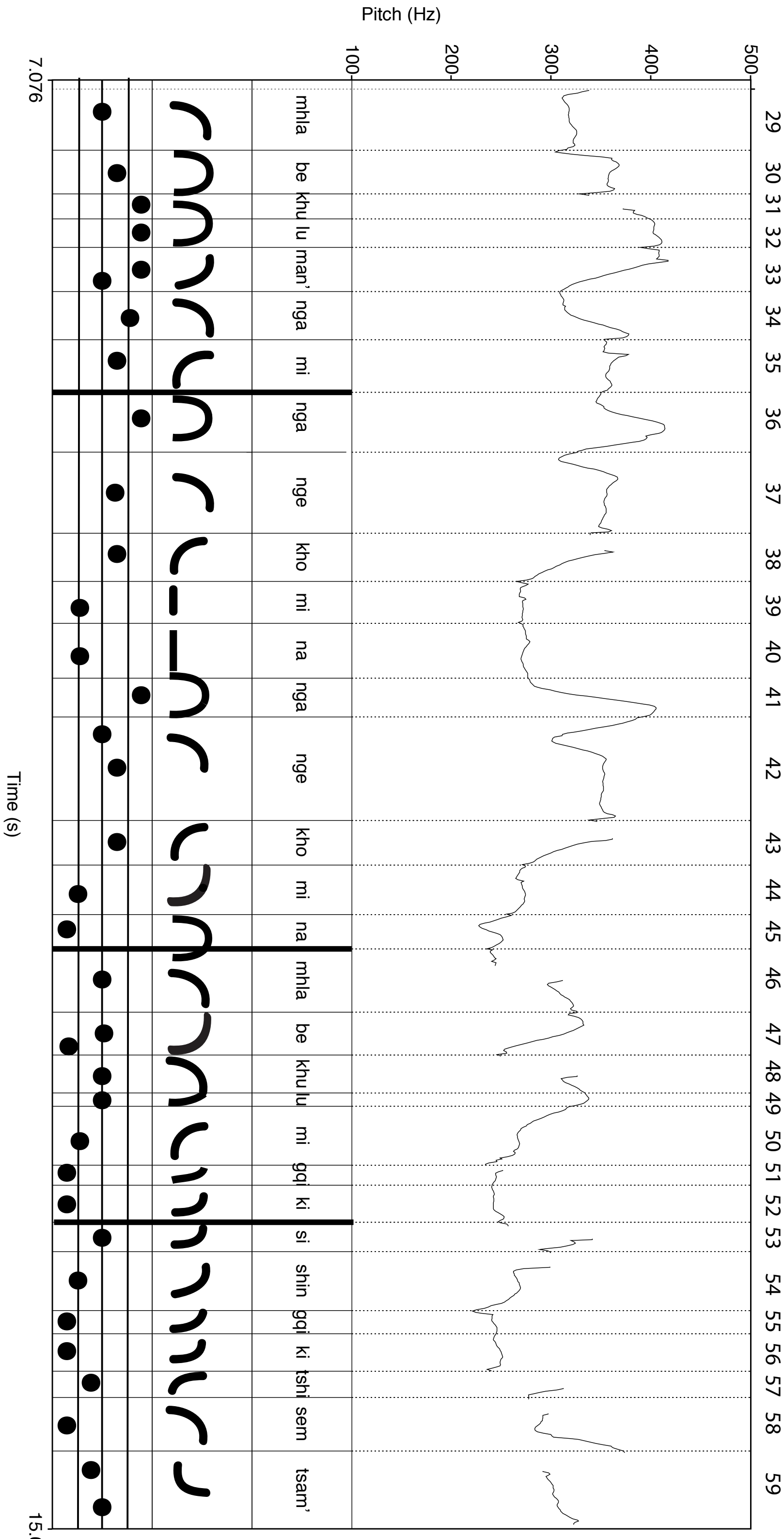
Most of the pitch categories are again consistent for the sung and spoken segments. The pitch contour of the spoken part appears to have more dramatic falls and rises. But it is in fact consistent when we scale the two graphs. All of the sung melodies extend over a much wider tessitura and usually begin at greater pitch height. The depressor action of stop consonants, nasals, and fricatives visible in the comparison is consistent with the other examples in this chapter.

Example 12.1: *Mhlabe* sung by Benzani Dladla (Part 1)

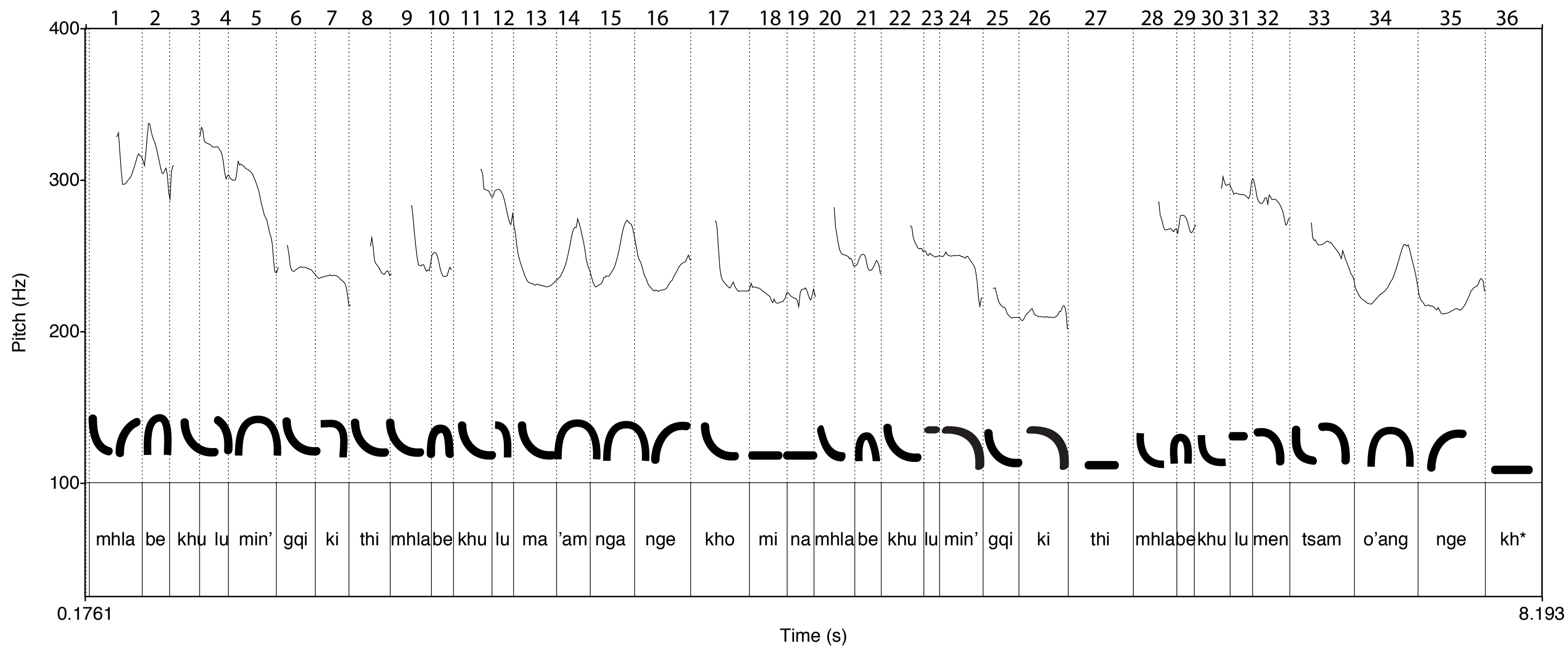


There are seven pitch targets over a range of less than an octave (i.e., a **heptatonic** system). Two main phrases and two subphrases consist as follows: A (1-14; 15-28), B (29-45; 46-59). These A and B sections may also be divided into couplets. This shows the underlying structure remains rooted in an interlocked call-and-response. Phrases A and B begin with a similar contour built around the word 'mhlabe.' At segment 15 the singer appears to make an error by carrying over the previous syllable. Instead of singing 'mhla-be' she sings 'tswa-be' thus repeating the syllable initial 'tswa' of the previous word. This shows how the context of words and syllables is an important factor in conditioning both articulatory and pitch gestures.

Example 12.1 Continued: *Mhlabe* sung by Benzani Dladla (Part 2)



Example 12.2: *Mhlabe* spoken by Benzani Dladla



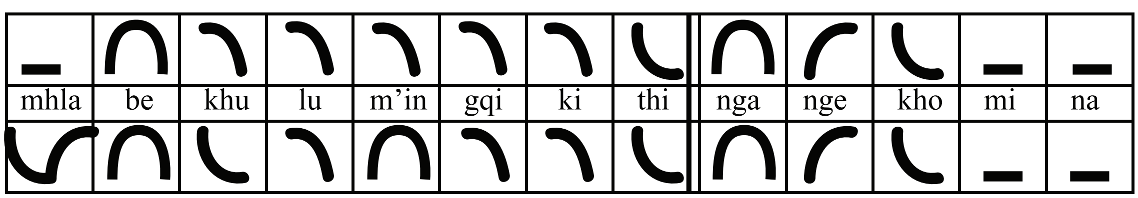
This example demonstrates the depressor action of click consonants in isiZulu. The alveolar click (‘gqi’) results in convex down-glides at segments 6 and 25. Other depressors in this example include the stop consonants ‘k’ (segments 3,7, 11, 17, 22, 26, and 30) and ‘t’ (segments 8, 33), and the fricative ‘hl’ (segments 1, 9, 20, 28). Fricatives, like sibilants, are always characterized by high to low frequency transitions and the associated pitch movements are always down-glides.

Lyric: *Mhlabe khuluma ingqikithi sisho ingqikithi isemtsamo. Ngangekho mina.*

The day they were talking about the secret of the ancestors. I was not there.

There are seven pitch targets over a range of less than an octave, framing a **heptatonic** system. Two main phrases and two subphrases consist as follows: A (1-14; 15-28), B (29-45; 46-59). Each phrase begins with a similar contour built around the word 'mhlabe.' At segment 15 the singer appears to make an error by carrying over the previous syllable. Instead of singing 'mhla-be' she sings 'tswa-be' thus repeating the syllable initial 'tswa' of the previous word. The context of words and syllables is clearly an important factor in establishing both articulatory and pitch gestures.

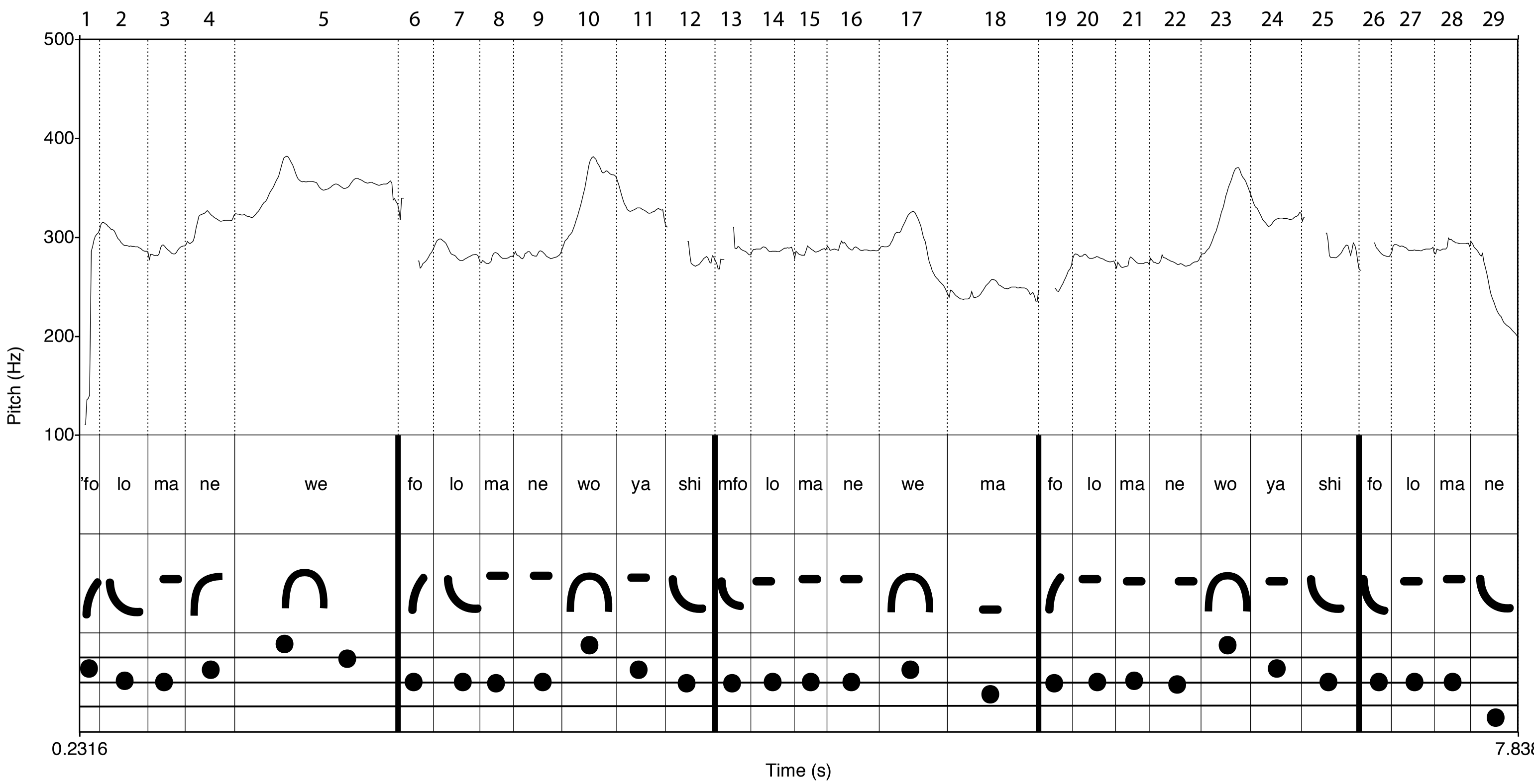
Example 12.3: *Mhlabe*, comparison of pitch categories for song and speech



This comparison shows that the pitch categories are nearly identical, and, more significantly, that pitch *movements* are the same for all segments.

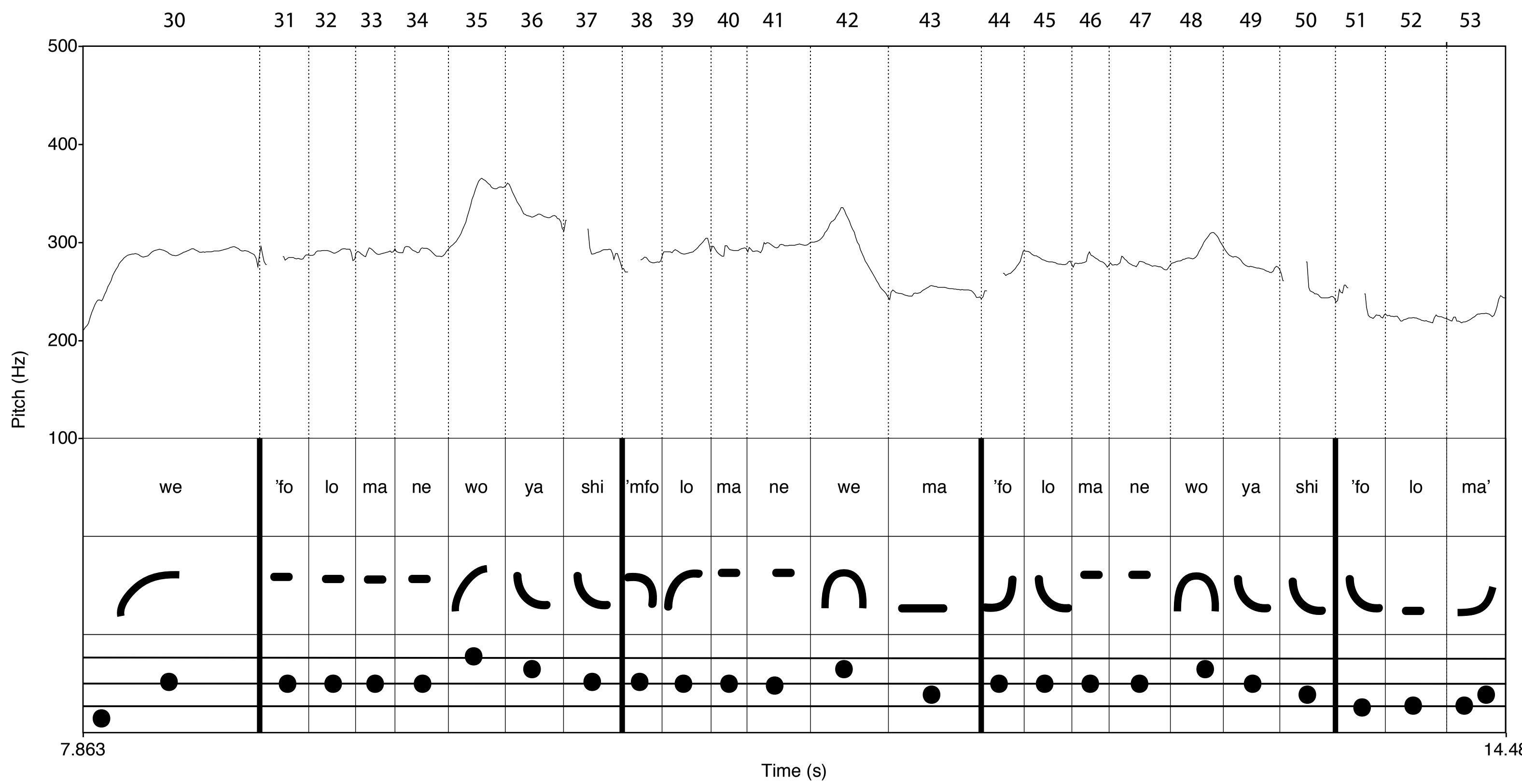
This example, more than the others, consists of several click consonants that have depressor functions. IsiZulu has three clicks (dental, alveolar and lateral) and all have a pitch-lowering function. We see the depressor action of alveolar clicks in segments 6 and 20. Stop consonants (11, 17), nasals (9, 19), and fricatives (13; 23) also control the pitch contour in this way.

Example 13.1: *Imfolomane* sung by Benzani Dladla (Part 1)

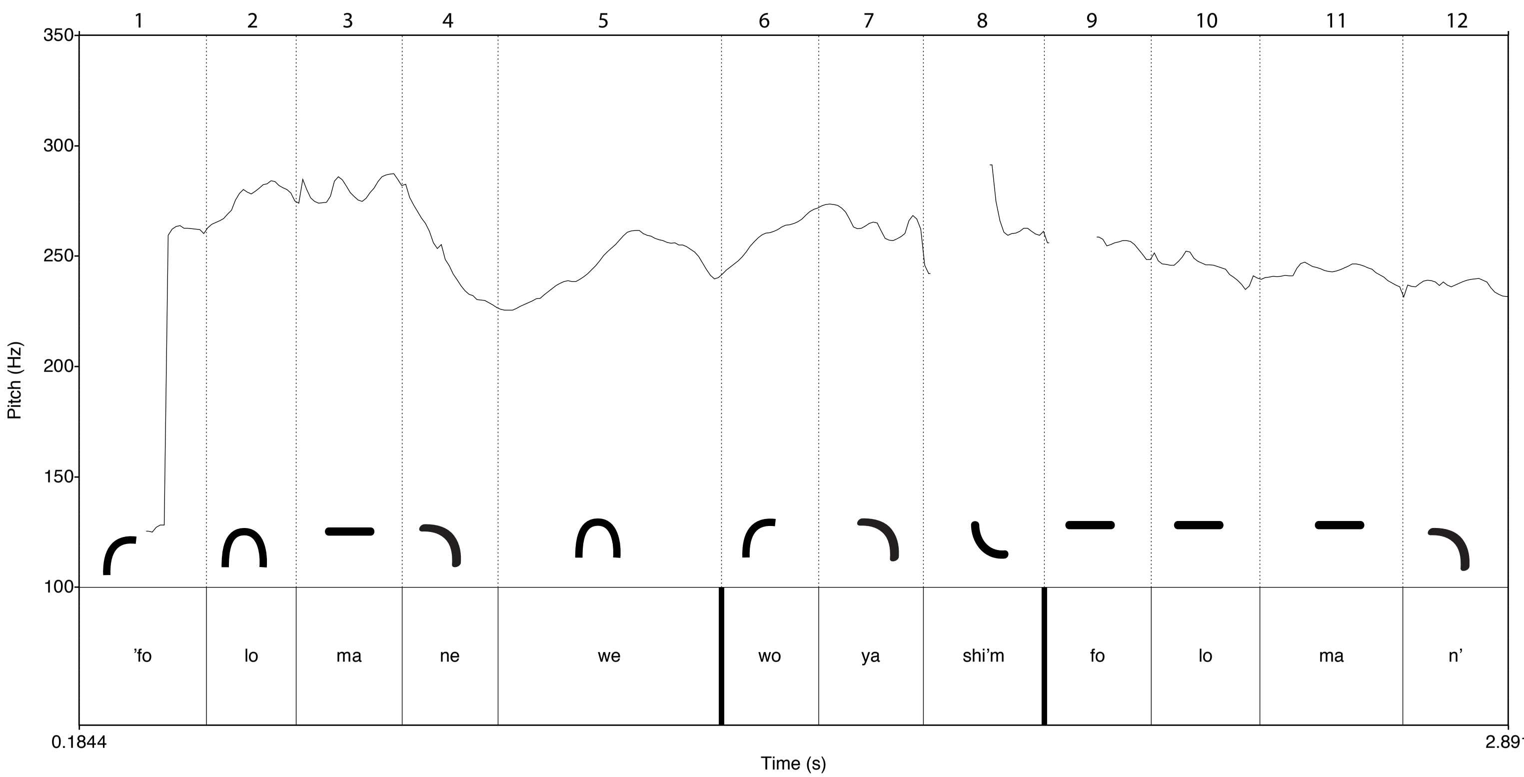


There are seven pitch targets in this song within a range of less than an octave (i.e., a **heptatonic** system). The phrase consists of couplets that alternate between: 'folomane we,' 'folomane woyashi,' and 'folomane we ma.' The pitch movements are very similar but the pitch targets vary. Notice the considerable variation in voice height between the High tone syllables 'we' (segments 5, 17, 30, and 42) and 'wo' (segments 10, 23, 35, and 48). This is despite the fact that the same syllables are used in the same phrase-structure as part of the same melody. This suggests that such variation is tolerated without a loss of identity in terms of the melodic percept. Listeners must well compensate for this by attending to the contour of the melody as a whole.

Example 13.1 Continued: *Imfolomane* sung by Benzani Dladla (Part 2)



Example 13.2: *Imfolomane* spoken by Benzani Dladla



The nasal depressor ‘ne’ results in a distinctive convex down-glide at segment 4. Notice how the sibilant ‘sh’ results in a similar down-glide at segment 8. This example is unusual in the number of level pitch categories employed. This may be explained by the relative paucity of depressor consonants in the lyric.

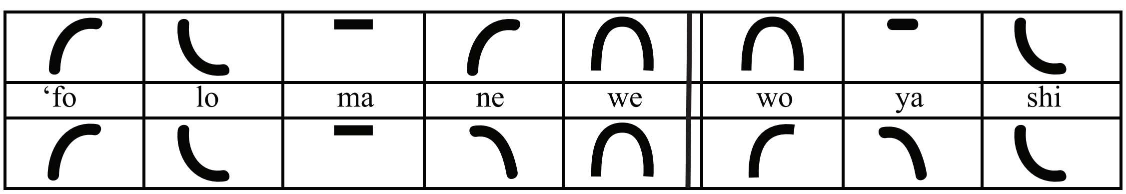
Lyric: *Woyasho imfolomane*

Foreman says [the *induna* (headman) has agreed for the ceremony to take place]

[The ‘foreman’ is the *induna* [headman] for a particular *isigodi* (district). In its more colloquial use it may also refer to a ‘foreman’ in charge of a gang of workers.]

There are seven pitch targets in this song that range of less than an octave, suggesting a **heptatonic** system. The phrase consists of couplets that alternate between: ‘folomane we,’ ‘folomane woyashi,’ and ‘folomane we ma.’ Within each of these there are very similar tonal patterns that reflect the spoken contour.

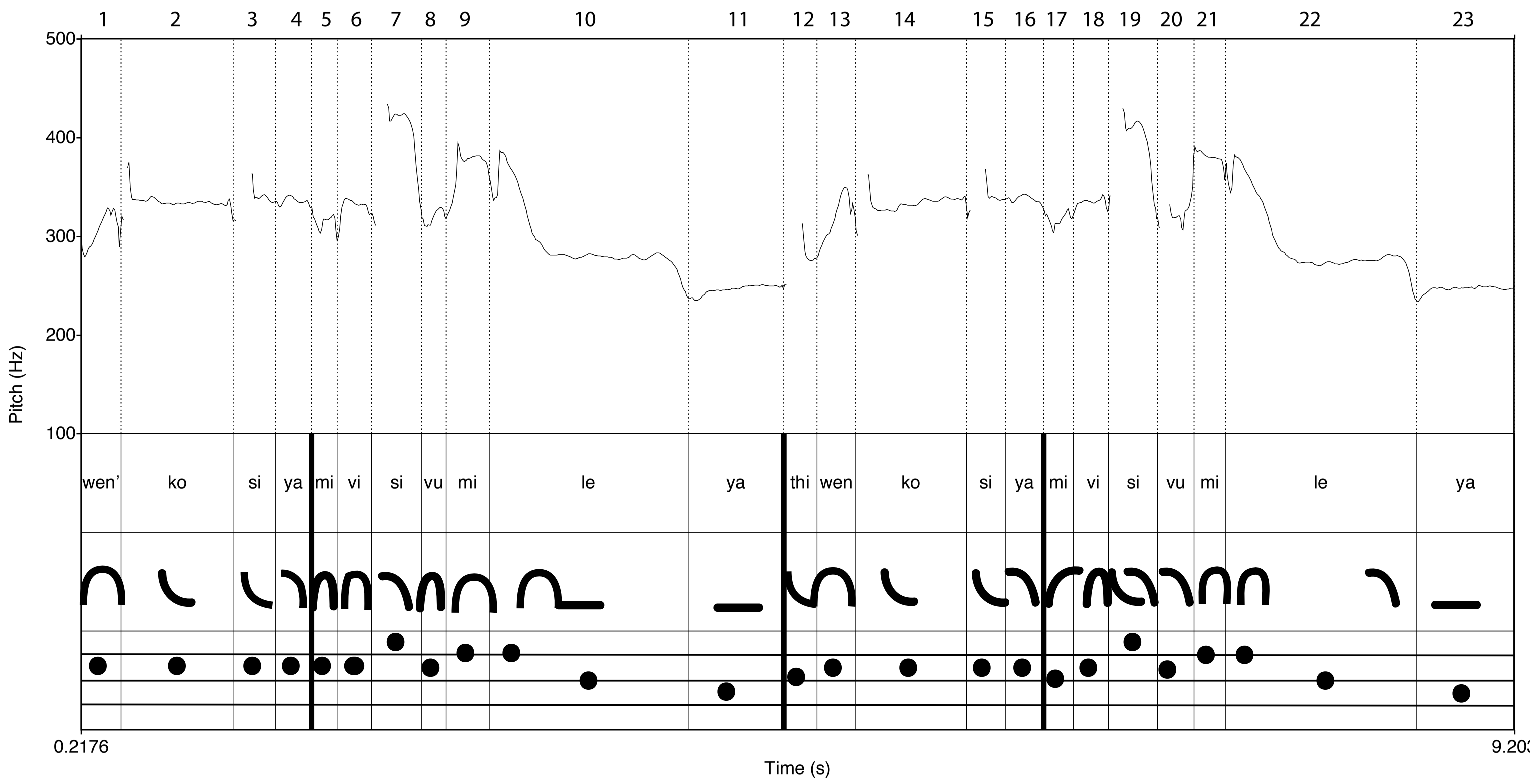
Example 13.3: *Imfolomane*, comparison of pitch categories for song and speech



The pitch categories for the sung and spoken portions of *Imfolomane* are similar but with one important exception: ‘ne’ is an up-glide in the sung version and a down-glide in the spoken version. This may be explained by the fact that in the sung version there is transition toward the next high pitch. This nullifies the effect of the nasal.

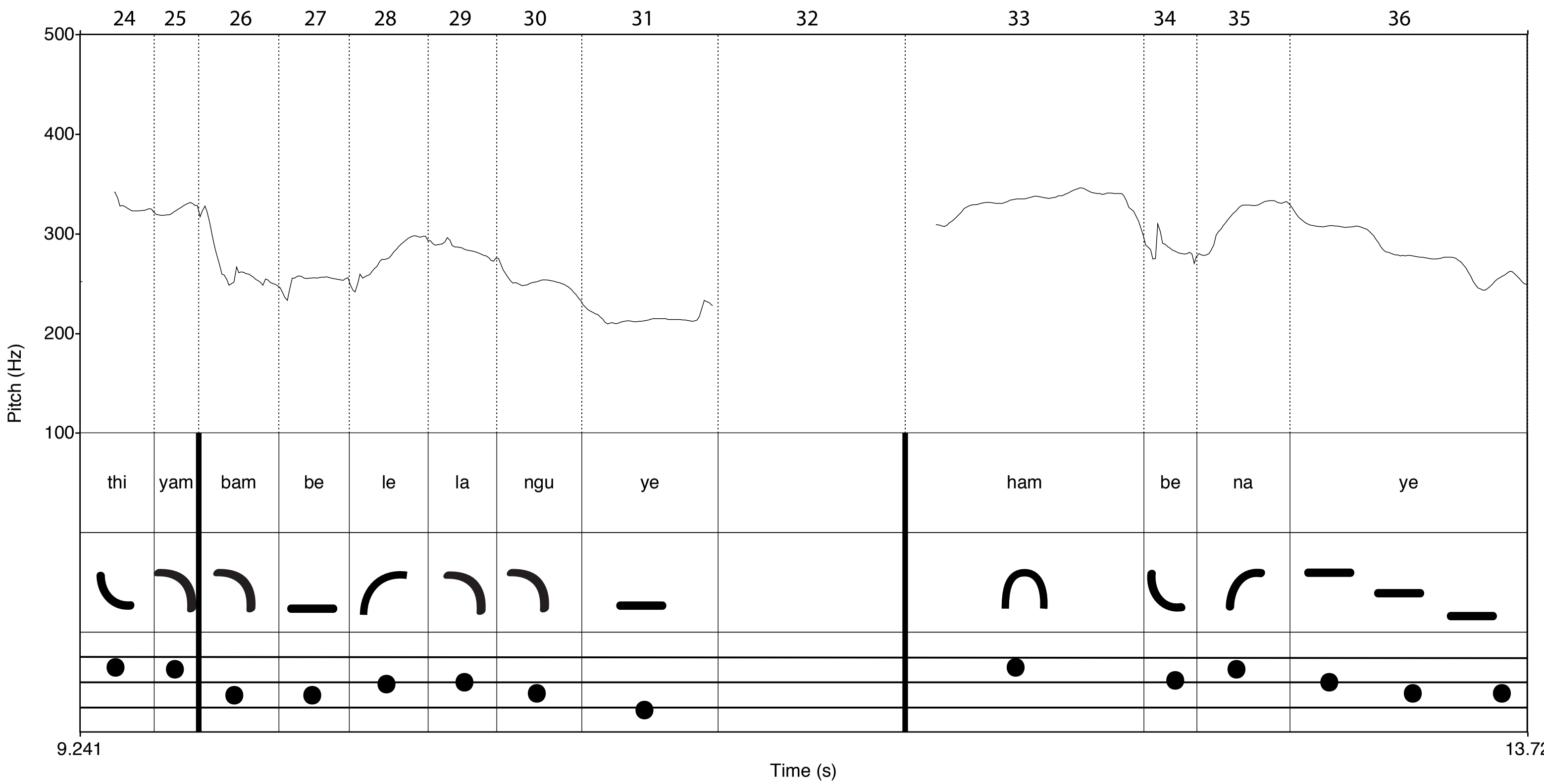
These examples show how articulatory (i.e. physical) features shape the pitch contour. The onsets to the fricatives in the ‘shi’ and ‘shi’m’ segments (12; 25) are steep concave down-glides. Similar down-glides occur on some nasals and laterals. These down-glides serve as important phonetic cues to these phonemes and hence to the semantic meaning of these phrases.

Example 14.1: *Inkosiyami* sung by Benzani Dladla (Part 1)

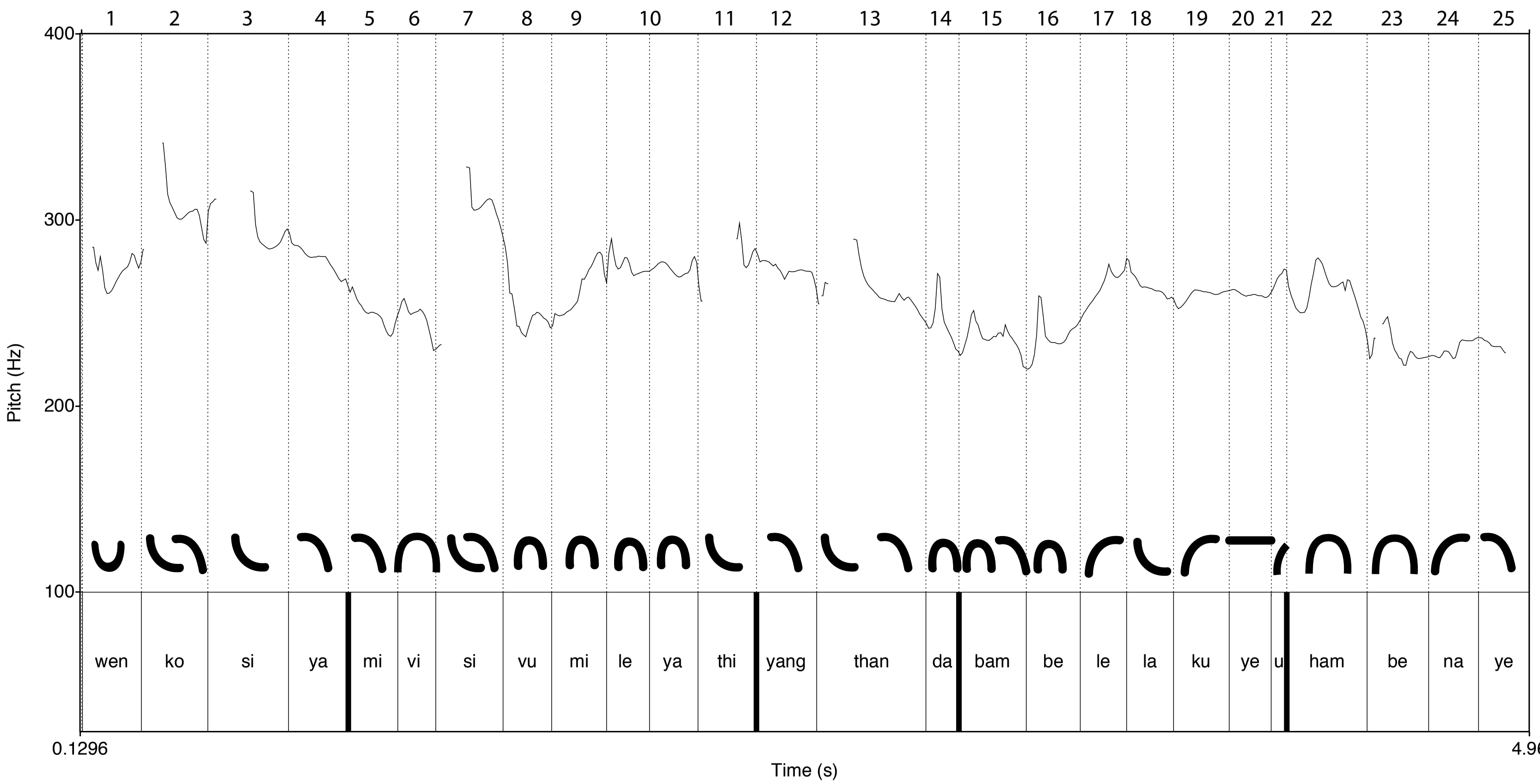


There are six pitch targets in this song and these extend over an octave range (i.e., a **pentatonic** system). The phrase structure is AAB (1-12; 13-24; 25 36). Segment 32 is left blank because there is a measured silence that balances the periodicity of the cycle. The lowest pitch target is reached at segment 31 with the final segment at 36 moving slightly higher and implying a return to A. This shows how the cyclic phrase structure tends to eschew closure in order to ensure continuity. Further evidence for this is found at segments 12 and 25 that give rhythmic impetus while acting as melodic bridges between sections. Notice the High, Mid, and Low tones (36).

Example 14.1 Continued: *Inkosiyami* sung by Benzani Dladla (Part 2)



Example 14.2: *Inkosiyami* spoken by Benzani Dladla



In this example the depressor function of stops (e.g., 2, 13, 11, 19), nasals (e.g., 12, 13, 15), and sibilants (e.g., 3, 7) is consistent with that of the previous examples. There is also noticeable downdrift and pitch-span reduction.

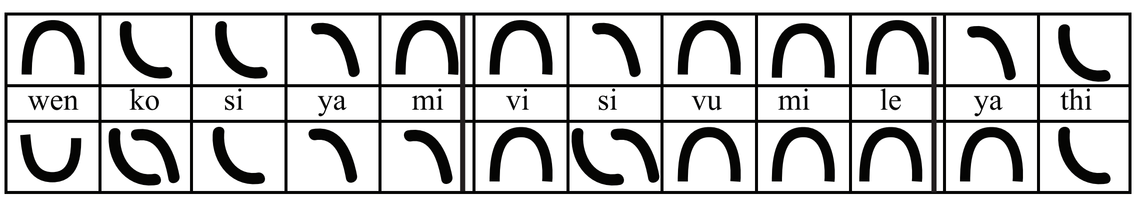
Lyric: *Wenkosiyama ibisivumile yathi iyamthanda. Bambelela kuye ihambe nayi.*

She has agreed that she loves him. So hold onto him and go with him.

[*Wenkosiyami* literally means 'oh lord' but is used as a common gesticulation or expressive effect]

There are six pitch targets in this song and these extend over an octave range outlining a **pentatonic** system. The phrase structure is AAB (1-12; 13-24; 25-36). Segment 32 is left blank because there is a measured silence that balances the periodicity of the cycle. The lowest pitch target is reached at segment 31 with the final segment at 36 reaching slightly higher and implying a return to A. This shows how cyclic phrase structure avoids closure to ensure continuity.

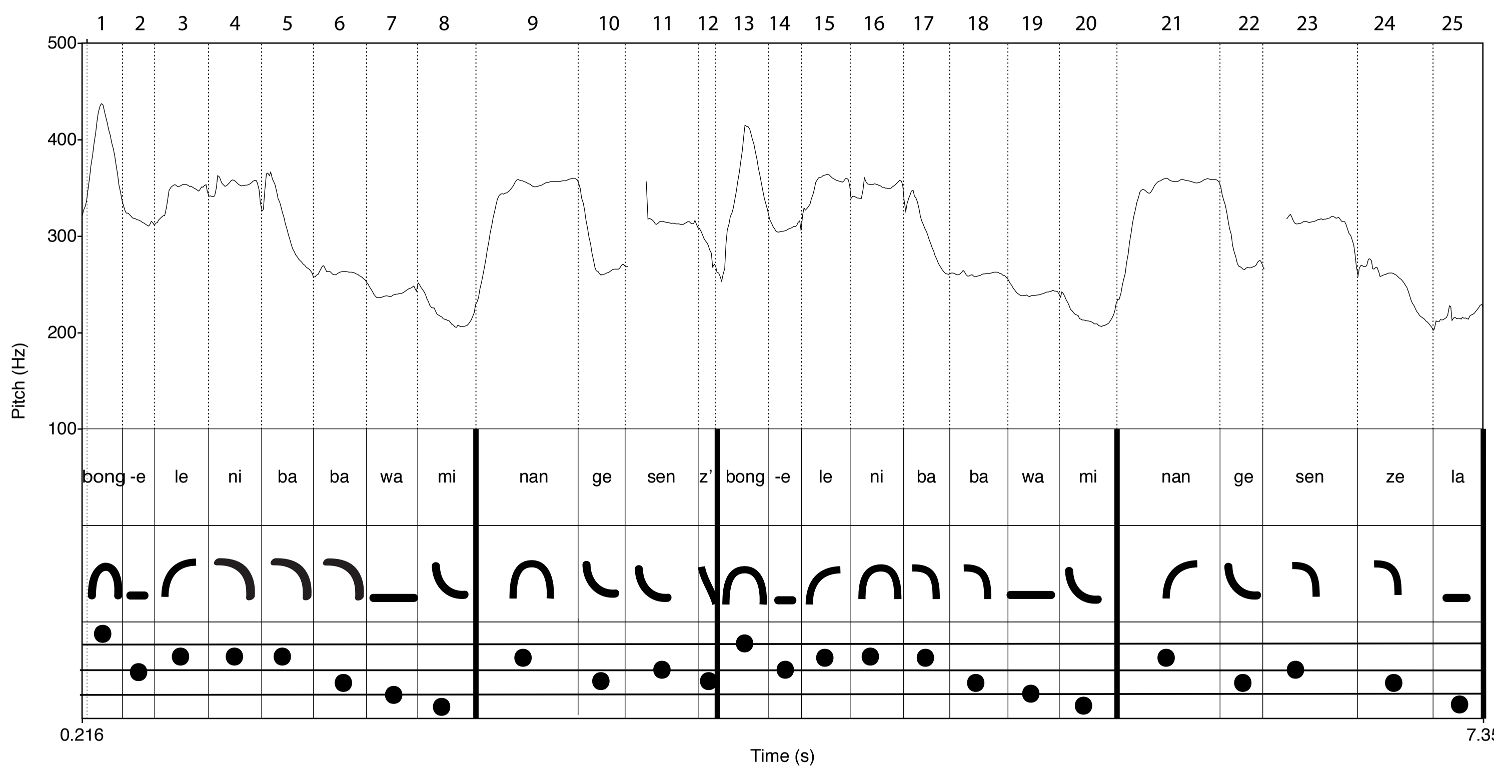
Example 14.3: *Inkosiyami*, comparison of pitch categories for song and speech



There are no significant deviations between the song and speech renditions of *Inkosiyami*. All pitch movements are the same, except for the very first segment ('wen') which is somewhat anomalous. However, given that there are no speech tone determinants on this segment the performance is underdetermined.

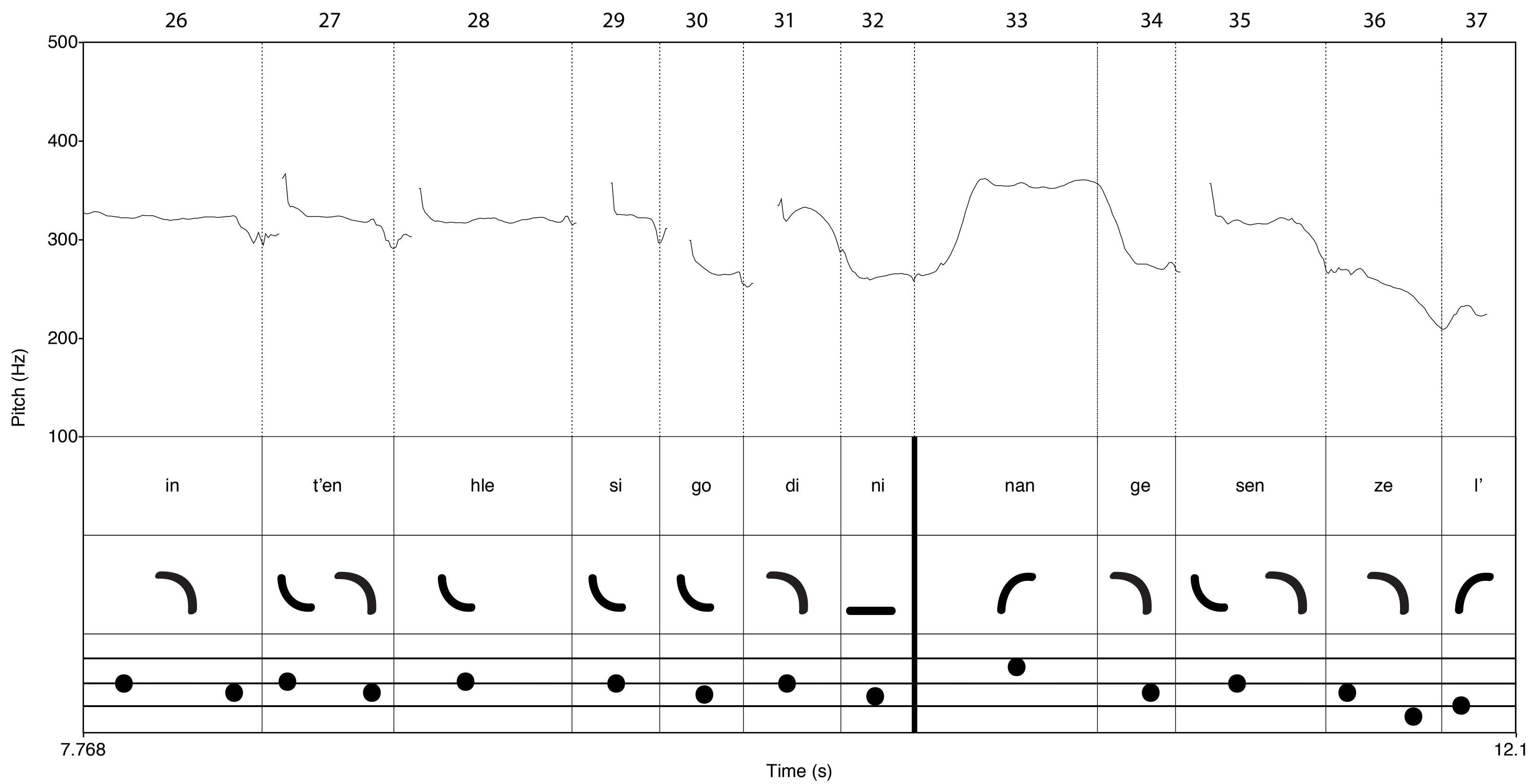
In this example the depressor function of stops, nasals, and fricatives is consistent but downdrift is less pronounced with each phrase begin at a similar pitch height. This suggests that downdrift is not determined by physical restrictions. Notice also how duration is used to articulate phrase structure. Longer note durations and *portamento* tend to accompany phrase-final pitch targets.

Example 15.1: *Baba Wami* sung by Benzani Dladla (Part 1)

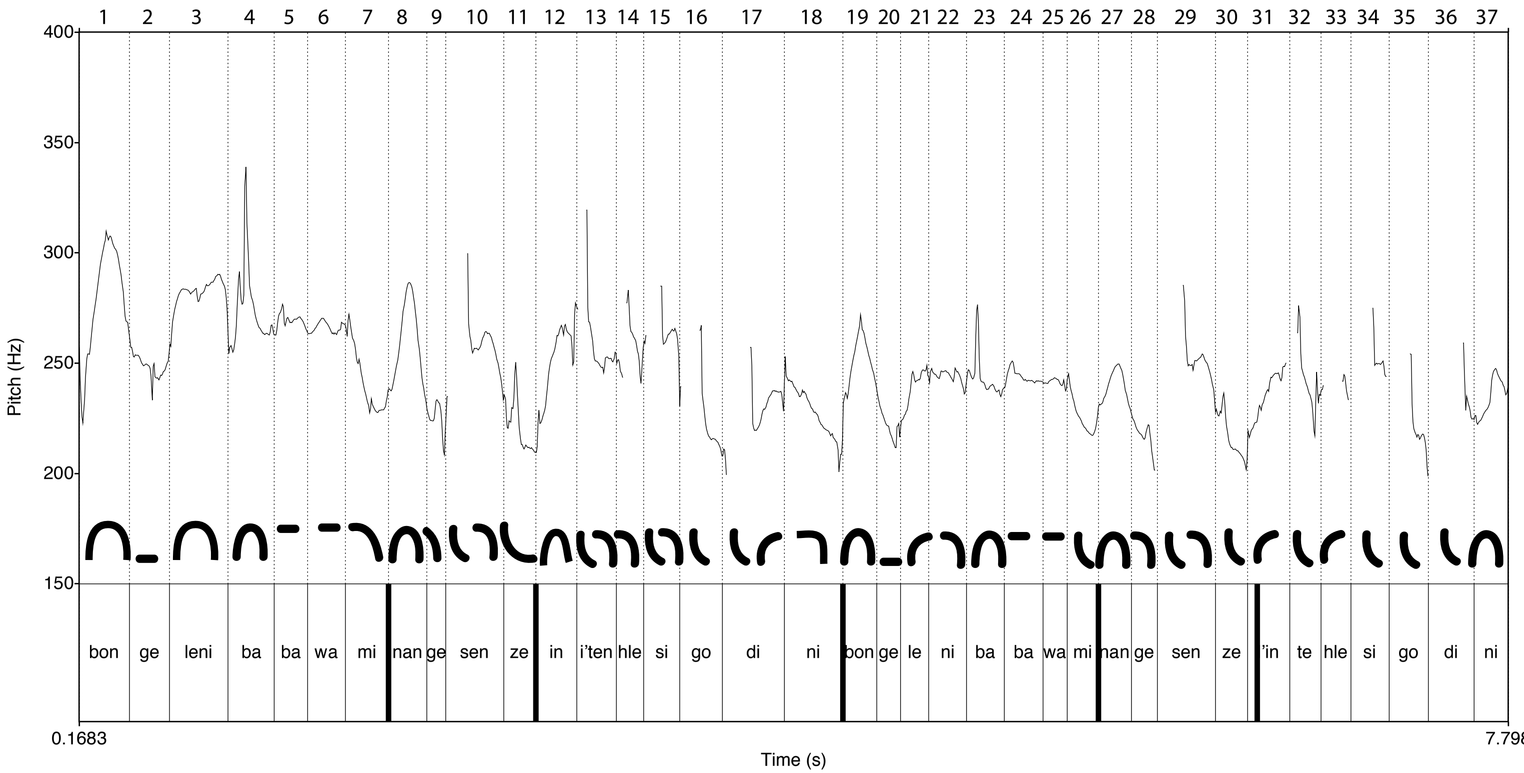


There are seven pitch targets spanning a range of more than an octave (i.e., a **hexatonic** system). The call (A) and response (B) structure underlying this song is structured as follows: The call (1-8) is followed by an abbreviated response (9-12) which is interjected with a return of the call (13-20) and the abbreviated response (21-25). We then have a new idea (C; 26-32) followed the the abbreviated response (33-37). This reduces to: **ABABCA**. Although there are technically three couplets in this structure, it is the CA group that provides both the tonal stability and also the impetus for the recurrence of the cycle. Once again, the interlocking call and response are the primary structural constituents and determine the shape of the melody.

Example 15.1 Continued: *Baba Wami* sung by Benzani Dladla (Part 2)



Example 15.2: *Baba Wami* spoken by Benzani Dladla



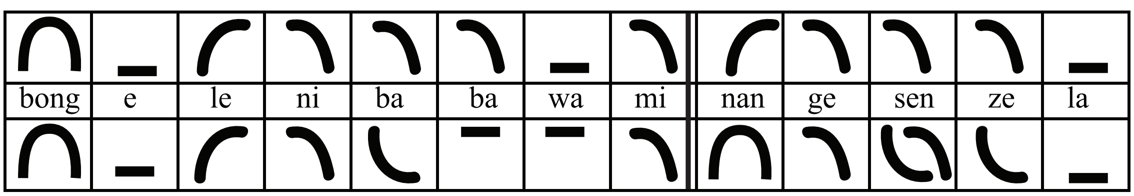
In this example we see many of the same pitch features as the previous examples: (1) depressor consonants including stops, fricatives and nasals, and (2) downdrift and reduced pitch-span. Concave contours in segments 1, 3, 19 and 27 may be explained by an initial up-glide to a High pitch target followed by a depressor consonant.

Lyric: *Ngibongeleni kubaba wami nangu esenzela into enhle esigodini*

I give thanks to my father for he is doing good things for us in this district.

There are seven pitch targets spanning a range of more than an octave outlining a **hexatonic** system. The call (A) and response (B) structure underlying this song is structured as follows: The call (1-8) is followed by an abbreviated response (9-12) which is interjected with a return of the call (13-20) and the abbreviated response (21-25). We then have a new idea (C; 26-32) followed the the abbreviated response (33-37). This reduces to: **ABABCA**. Although there are technically three couplets in this structure, it is the CA group that provides both the tonal stability and also the impetus for the recurrence of the cycle.

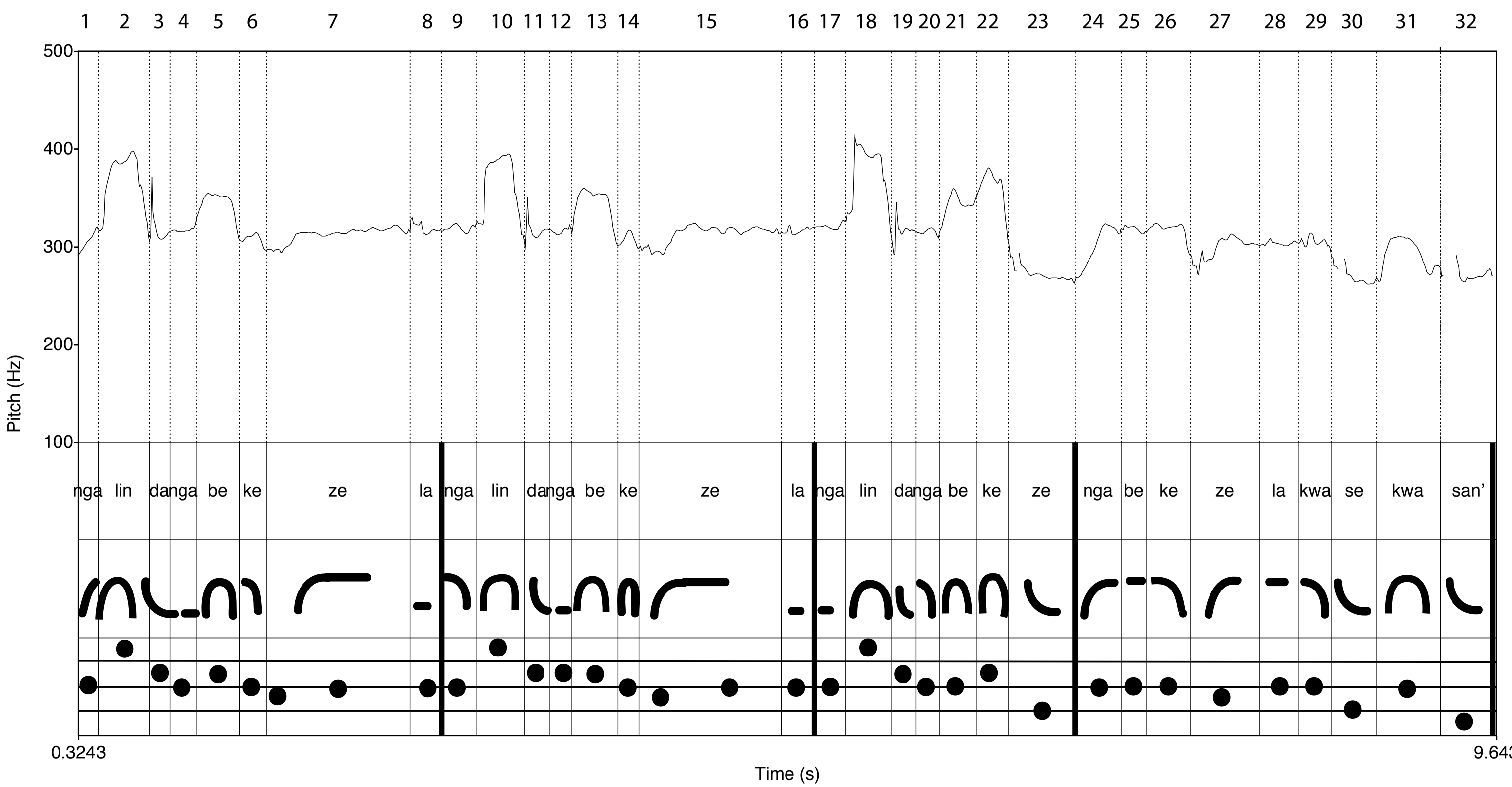
Example 15.3: *Baba Wami*, comparison of pitch categories for song and speech



The pitch categories are similar for both renditions of *Baba Wami* (song and speech), and the pitch movements are the same. Again, the only differences have to do with the presence or absence of glides, and these are context-specific.

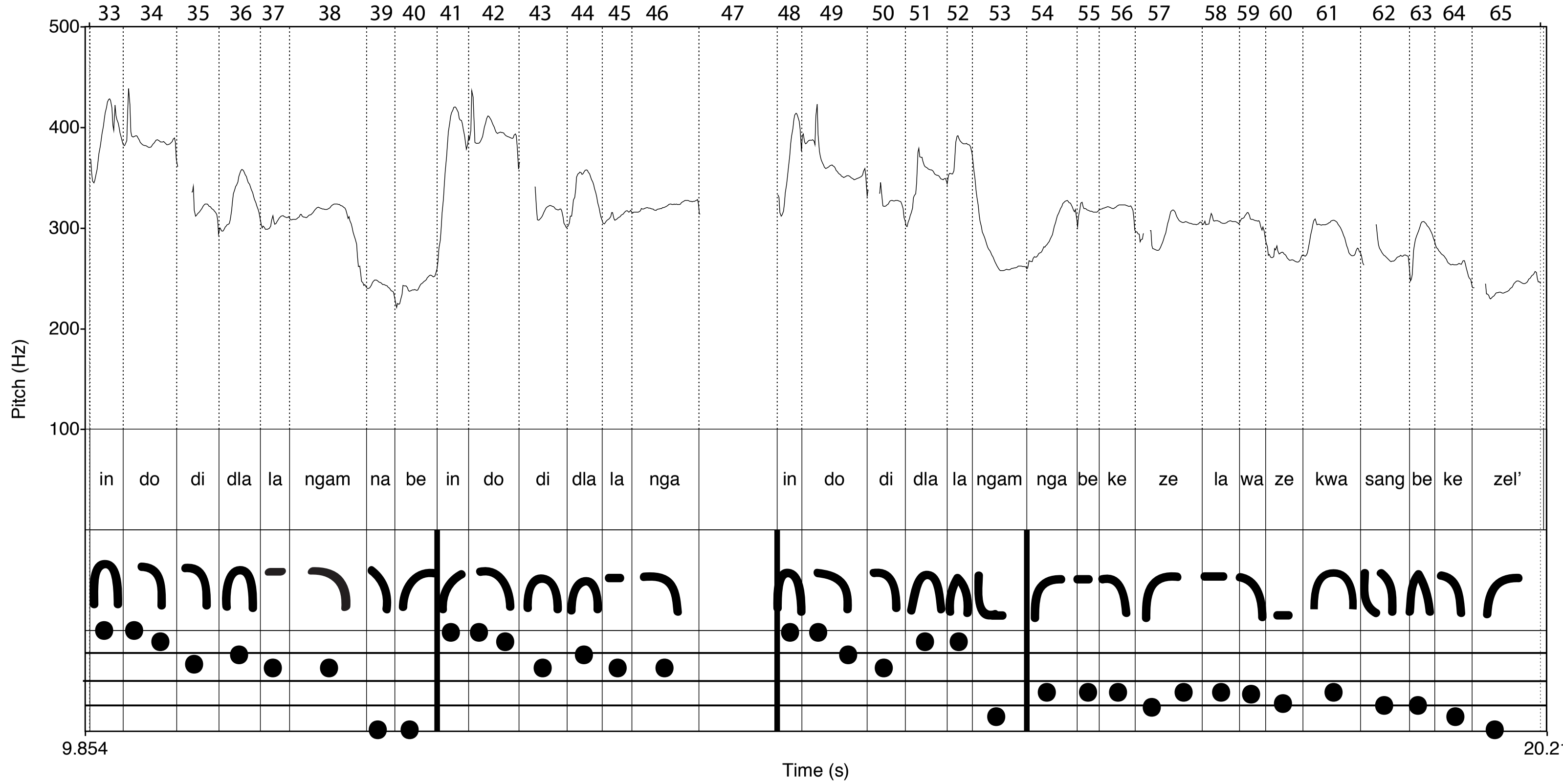
This example displays one of the more complex pitch systems. Notice how few level pitch categories there are. Given that speech-tone is contrastive this means that contours and glides must fulfill this function. There is still a high degree of overlap between the song and speech categories with stops, nasals and fricatives resulting in similar depressor features.

Example 16.1: *Ngabekezela* sung by Benzani Dladla (Part 1)

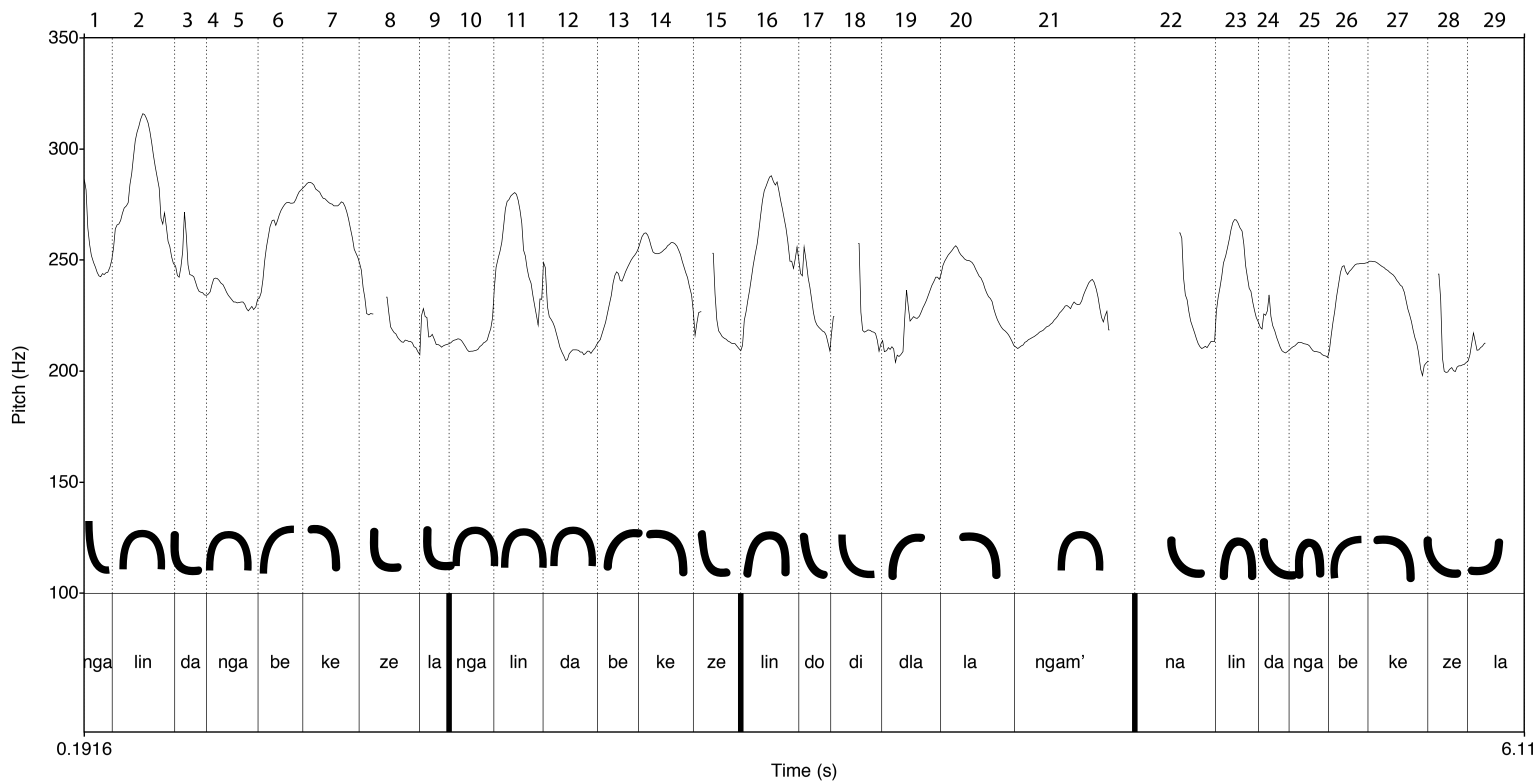


There are nine pitch targets over more than an octave range (i.e., a **heptatonic** system). The phrase structure is $AAA^1BB^1B^2A^2$ (1-8; 9-16; 17-23; 24-32; 33-40; 41-47; 48-53; 54-65). This is a clear example of the use of *variants* of the same phrase that are concatenated and reshuffled for melodic purposes even though the semantic meaning of these scrambled phrases is more or less the same. The periodicity of the cycle seems to be unshakeable with all other pitch elements structured according to its underlying momentum.

Example 16.1 Continued: *Ngabekezela* sung by Benzani Dladla (Part 2)



Example 16.2: *Ngabekezela* spoken by Benzani Dladla



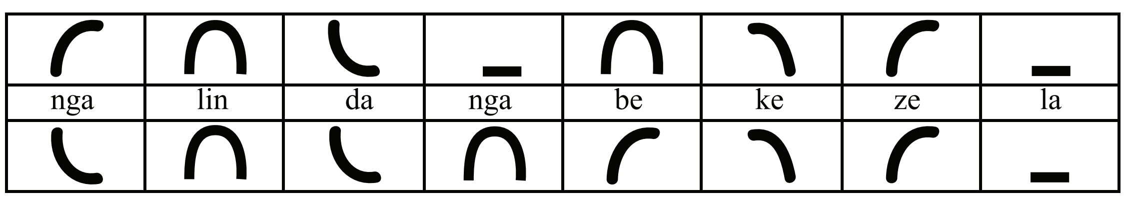
There are many peaks and troughs in this example. These are conditioned by High tones and by the action of depressor consonants, including: stops, fricatives, sibilants, and nasals. There is a gradual reduction in the pitch-span with the ceiling lowered over the course of the utterance. The High pitches are lower toward the end while the Low tones remain relatively constant in pitch height (from segment 9 onward).

Lyric: *Ngalinda ngabekezela kwaze kwasa ngibekezele. Nansi indoda idlala lami. Yangishela ngoba ngithanda. Isingibhangqa ninondindwa.*

I have waited patiently until the morning, patiently. Here this man of mine is playing. He proposed to me because he loves me. He is having an affair with this slut.

There are nine pitch targets over more than an octave range suggesting a **heptatonic** system. The phrase structure is $AAA^1BB^1B^2A^2$ (1-8; 9-16; 17-23; 24-32; 33-40; 41-47; 48-53; 54-65). This is a clear example of the use of *variants* that are concatenated and reshuffled for melodic purposes even though the semantic meaning of these scrambled phrases is more or less the same. The periodicity of the cycle seems to be unshakeable with all other melodic elements structured according to its momentum.

Example 16.3: *Ngabekezela*, comparison of pitch categories for song and speech



The pitch categories for spoken and sung versions of *Ngabekezela* are very similar. Only the first segment ('nga') requires explanation due to the contrast in pitch movement. In the spoken version there is a convex down-glide. This is expected because of the depressor action of the stop consonant. However, in the sung version the singer is aiming for an initial High pitch target to begin the melody, and so melodic considerations outweigh speech-tone determinants.

The song and speech contours diverge only slightly due to concatenation of syllables in and, in some cases, the interpretation of contours as level tones. Stops, nasals, and fricatives perform similar depressor functions.

DISCUSSION

Each of these analyses offers insight into the workings of Zulu song prosody as a multi-dimensional and complex adaptive system. The evidence provided here does not support the notion of an underlying harmonic structure. There are three main reasons for this: First is that the pitch *contours* of song and speech are very similar. If we compare the upper systems from each of the song and speech graphs we see how that is a degree of parallelism that we would not expect to find if song had a different cognitive foundation than speech. This point is supported by the comparison of pitch *categories* in the third graph of each analysis. The *Umthetho* example shows that different singers will use different pitch spans for the same song. This means that the pitch categories they employ cannot be discrete even if they are categorical. The identity of a song should not, therefore, be based entirely on its set of 'discrete pitches,' but rather on the set of pitch categories and contours that define its *shape*. Another point about these categories is that gradient features are far more common than are level pitches we might find in a staff notation transcription.

A second main reason to be skeptical of harmonic explanations is that there is virtually no harmony present in choral dance music; one must infer an implied harmonic structure. The antiphonal, cyclic structures characteristic of dance and ceremonial songs are not in any case conducive to harmonization because of their overlapping phrase structures. In the few genres where harmonized choral singing does take place—such as in *amahubo* (anthems)—it seems that speech-tone determinants largely restrict chords to octaves and open fifths and fourths. Even in *isicathamiya*, a genre that grew out of mission hymnody, homophonic textures and harmonized settings are rare.

The third reason is perhaps the most compelling: the gradual narrowing of the *pitch-span* of each cycle is evidence for the perceptual primacy of contour.[259] If we measure pitch-span using a *baseline* and a *topline* of pitch targets (such as the bottom rung of the above pitch analysis graphs indicate), then we have a clear floor and ceiling for the pitch-span analysis. Analysis of the width of this pitch-span shows two things. First, there is an average descent or declination in pitch. Second, there is also the *narrowing* of the pitch-span—or the average distance between baseline and topline—over the course of each song. This narrowing can only take place if the information encoded in discrete pitch categories is but one of several cues for the listener; otherwise the narrowing would result in an unrecognizable melody. This aspect of the song impels us to use more flexible schemata to model melody as a networked phenomenon, a product of multiple pitch categories all of which are flexibly encoded—that is, a gradient rather than a discrete system.

Analysis of these ten *memulo* songs and their spoken texts reveals that the articulation of consonants plays a decisive role in accounting for the shape and contour of melody. These findings replicate those of Rycroft and other linguists in demonstrating that stop consonants, nasals and fricatives all have a pitch lowering function in Zulu song and

---

[259] See John Laver, *Principles of Phonetics* (Cambridge: Cambridge University Press, 1994): 459-460.

speech. To these efects we may add another linguistic factor that has not been remarked upon in this literature: this is the effect of *vowel height*. It is clear from each of these analyses that vowels act as *pitch targets*. Not only that, but high vowels and low vowels tend to carry appropriately valenced pitch targets. Vowels are pitch targets because they are always voiced and stable and because they do not shift unless acted upon by a consonant articulation. My hypothesis here is that front vowels are more often accorded High tones while back vowels tend to take Low tones. This observation requires further investigation and experiment but it points to a physical restriction on pitch production that is similar to the effect of depressor consonants.

Each example included a graph of pitch targets that had been extracted from the overall contour. One goal of this chapter was to establish the nature of these pitch targets and their organization in melodic structure. Although several scale systems were alluded to in the analyses—including pentatonic, heptatonic, and hexatonic orderings—no attempt was made to establish a hierarchy of pitch relations within these. Most theories of tonality describe a functional hierarchy for scale systems. But the cyclic structure and gradient pitch categories of Zulu song prosody tend not to conform to such hierarchies. These pitch target analyses are perhaps better used for studying *pitch-span* because they show the diversity of possible pitch spaces and the extent to which even a single singer in a single repertory will use widely divergent set.[260]

---

[260] Note that the use of the term ‘pitch-span’ in phonetics is very different to its usage in generative music theory. I use it only as a reference to the difference between the ceiling and floor of a particular passage.

FINDINGS

Based on this evidence we may offer several preliminary findings for Zulu song prosody:

1. Pitch is not consistently discrete. Pitch targets are rarely self-evident categories when analyzing fundamental frequency tracks.
2. Level pitches are categorical but so are pitch gradients. They are not simply 'intermediate' to high or low level tones. The evidence presented here suggests that gradient pitch categories are more common than level pitch, and that the latter tend to signal phrase boundaries.
3. Pitch contours act as phonetic cues that are vital to unambiguous communication of semantic information. The action of depressor consonants is one clear case in point. If we flatten out these contours, including the microprosodic details of onsets and codas, then we ignore perceptually salient information for distinguishing phonemes.
4. Pitch contours are determined primarily by linguistic rather than 'musical' considerations, with a few exceptions. At phrase and cycle endings we sometimes find discrepancies, as Rycroft remarked: "Usually a particular note serves as a melodic *finalis*. This note may have to be approached by way of some accepted terminal or cadential interval progression and the resultant melodic sequence tends to take precedence over speech-tone requirements."[261] While it is true that phrases tend to end at low points in the contour, this is not the case for all

[261] Rycroft, *Indigenous* Music, 312.

utterances. Furthermore, the comparison of song and speech contours shows that this phenomenon is consistent. It may simply be a product of downdrift.

5. Pitch contours in song and speech are very similar. The rise and fall of contours and pitch categories are nearly identical. Differences arise in the overall pitch height with speech almost always occupying a lower pitch range than song. This is probably the result of the increase in pressure required to project the singing voice.
6. Pitch is conditioned by language- and culture-specific elements more than by acoustic properties of the auditory system, including the overtone series. We should be very cautious in adopting a tonal framework founded on the harmonic series. This was pointed out as early as 1885 by Alexander Ellis in his cross-cultural analysis of scale systems.[262]
7. Pitch is conditioned by depressor consonants that include nasals, fricatives, sibilants and stops. These are the product of articulatory mechanisms and serve as phonetic cues, but their insertion in a phrase or cycle may also be used for melodic purposes. That is, phonemes (consonants and vowels) with particular tonal qualities may be used to generate pitch effects expressly for musical rather than semantic purposes.

In all cases we see how melody is the product of a multi-dimensional prosodic structure that include suprasegmental features of speech tone and intonation as well as a complex

[262] Alexander Ellis, "On The Musical Scales of Various Nations," *Journal of the Society of the Arts* 33 (1885): 485-517.

paradigmatic pitch structure. Finally, the pitch categories introduced in the previous chapter, and applied in this one, provide a useful theoretical framework for modeling prosodic structure.

CONCLUSION

This chapter set out to establish the basic features of Zulu song prosody. Parts 1 and 2 included a review of musical and linguistic determinants of pitch patterning in song. I argued against the notion of an underlying harmonic system and in favor of a linear design rooted in the phonological treatment of pitch categories that show gradience and variation in their realization. The *memulo* songs studied in Part 3 provide empirical support for this theory and implement a new method for the phonetic analysis of song melody using Praat. The findings of this chapter, based on these analyses, suggest that we should conceive of melody not as a discrete combinatorial system but as a *gradient* system that is nevertheless combinatorial in its own right. It seems that a large part of our perception and cognition of melody in song is the result of our inferring categories that are both discrete and gradient, local and global; categories that cannot, in short, be explained by computational models. The gestures inherent in pitch contours—both real and implied—provide one vital component in our perception of melody. In studying these gestures and their origins in human articulatory movements we begin to see how speech and song are embodied phenomena with fluid boundaries. And in taking a cross-cultural perspective we find that what Westerners consider categorical distinctions are not evident in other world music cultures. Categories are conventions that are culturally learned.

Together, this evidence challenges long-held assumptions about the nature of tonality, discrete pitch and combinatoriality, hierarchical structure and contour in music theory. Phonetic studies that employ the sorts of analytical rubrics developed here offer an important alternative for investigating melody as prosody without making assumptions about the very categories of perception and production that we seek to investigate.

CHAPTER 5

Zulu, KwaZulu, and IsiZulu: On fieldwork

Ethnomusicologists record and render their interactions with cultural others primarily as outsiders, whether or not they are members of the culture. Careful self-reflexive gestures choreograph a literary document, since the discipline is still very much one of writing culture.[263] Fieldwork emphasizes ethnography as *experience* and foregrounds social aspects over and above the 'music itself'.[264] Since the 1970s, recordings, transcriptions and analyses gradually receded in importance with the decline of comparative and systematic musicology, and for many ethnomusicologists close analysis became a minor or even peripheral concern. Today this situation is again changing with a recent return to 'analytical studies in world music.' But there remains a great deal of resistance to these approaches, as I aim to show in Part 1 of this chapter.

My study takes a very different approach to fieldwork than 'cultural ethnomusicology' because I locate my work within a cognitive science paradigm. Cultural immersion was not the goal of this study—and it was indeed impossible, given the multi-sited program—but a discussion of the social contexts and conditions of my fieldwork is nevertheless

---

[263] James Clifford and George E. Marcus, eds., *Writing Culture: the poetics and politics of ethnography* (Berkeley: University of California Press, 1986).

[264] "[F]ieldwork is experience, and the experience of people making music is at the core of ethnomusicological method and theory." Timothy J. Cooley and Gregory Barz, "Casting Shadows: Fieldwork Is Dead! Long Live Fieldwork!" In *Shadows in the Field: New Perspectives for Fieldwork in Ethnomusicology*, Second Edition, ed. Barz and Cooley (New York: Oxford, 2008): 14.

important. Understanding the contexts in which musical and linguistic performances are made and broadcast is vital if we are to study their communicative functions. Fieldwork is a social process and so—no matter what the research goals—it is important to recognize both the impact of researches on communities and how fieldwork conditions shape the final product. This is especially important for this project because I relied entirely on ethnographic recordings made over the course of many months of fieldwork.

In Part 1 of this chapter I discuss ethnomusicology's troubled relationship with cognitive science before discussing my own approach to fieldwork in Part 2. This section of the chapter also provides contextual and historical information on Zulu music and culture that complements the case study outlined in Chapter 4. Finally, in Part 3 I reflect on the practical and intellectual challenges of conducting music cognition research in the field as opposed to in the laboratory.

PART 1: ETHNOMUSICOLOGY AND COGNITIVE SCIENCE

For some time now ethnomusicologists have recognized the growing influence of the cognitive and hard sciences in the humanities and social sciences. Most have opposed such developments. Judith Becker, reflecting on the history of the disciplines, writes that "[e]mpirical research fell out of favor in the discipline of ethnomusicology with the general discrediting of […] Comparative Musicology, and the concomitant rise of Cultural Anthropology as a signature discipline in the mid-twentieth century. Ethnomusicology followed this general trend and became almost exclusively Cultural

Ethnomusicology."[265] In his response to Becker in the *Journal of Ethnomusicology* Jeff Todd Titon writes that, "the poststructuralist critique of science has exerted a powerful influence on ethnomusicology in recent decades. As in anthropology, an interpretative turn came to characterize work in the discipline: interpretation rather than analysis, understanding rather than explanation, and individual experience as the site of culture."[266] What both scholars acknowledge is a widening gulf between methods of social and cultural anthropology and the interdisciplinary agendas of cognitive science. Indeed, the introduction of neuroscientific and biocultural perspectives that stress such topics as universals and evolutionary theory have met with skepticism and hostility in ethnomusicology, with a number of important exceptions to be discussed below. Understanding music across cultures has become a peripheral concern. Instead it is immersion in cultural practices that is prized in the 'relativist' position Titon describes.

Ethnomusicology's break with science was gradual, but began to solidify in the 1970s as ethnographic fieldwork was entrenched as the core method of inquiry. Not long before that the founding generation of ethnomusicologists had regarded their work as an extension of the sciences. In 1964, Alan Merriam, perhaps the most influential proponent of the anthropological perspective in ethnomusicology, described the young discipline as "sciencing about music";[267] in 1971 Mantle Hood claimed that, "comparative studies at

---

[265] Judith Becker, "Ethnomusicology and Empiricism in the Twenty-First Century," *Ethnomusicology* 53, no. 3 (2009): 478.
[266] Jeff Todd Titon, "Ecology, Phenomenology, and Biocultural Thinking: A Response to Judith Becker," *Ethnomusicology* 53, No. 3 (2009): 503.
[267] Alan Merriam, *The Anthropology of Music* (Evanston, IL: Northwestern University Press, 1964): 25.

various levels of concentration can be identified among the important objectives of the ethnomusicologist."[268] Bruno Nettl, also one of the pioneers in the 1950s, has recognized the paradox that has arisen in more recent work:

> In music research, the concept of comparison has, over the last fifty years, come to have a questionable and even unsavory reputation; nevertheless, a large proportion of the most significant research depends on comparison of one sort or another for its identity and effectiveness. Some of the concepts of greatest currency in recent scholarship – e.g., "world music," universals, evolution of music, diasporas – require, for their development, some kind of comparative approach.[269]

This "unsavory reputation" is the product of many factors, but the rise of Cultural Anthropology and its resistance to scientific method are crucial among them.

> [S]ince the "crisis" in the 1970s over representation, authority, and objectivity, and over ethnography's complicity with colonialism and empire, ethnographers in North America have not been reductionist: as a rule we are self-reflexive, interpretative rather than analytical, subjectively present in our work and writing, and intersubjective in our claims for understanding. We espouse cultural relativism, we believe reality is humanly and socially constructed and maintained, we privilege individual experience, we believe in human agency, and we are interested in giving back to the people in the communities we study. In a word, we are humanists – mostly.[270]

---

[268] Mantle Hood, *The Ethnomusicologist* (Kent, OH: McGraw-Hill, 1971): 342.
[269] Bruno Nettl, "Comparative Study and Comparative Musicology: Comments on Disciplinary History," in *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology*, ed. Rolf Bader, Christiane Neuhaus, and Ulrich Morgernstern (Frankfurt: Peter Lang, 2010): 295.
[270] Titon, "Ecology, Phenomenology and Biocultural Thinking," 503.

Titon describes questions of relativism, subjectivity, and individualism as incommensurable with the 'reductionist' mode of scientific inquiry associated with comparative musicology. Post-Geertzian ethnomusicology has distanced itself from the methodological premises of comparative musicology, thus alienating important historical figures such as Von Hornbostel, Sachs, Stumpf, and Kolinski and their substantial contributions to analytical studies of world music.[271] Nettl describes how the reflexive turn resulted in a new outlook for ethnomusicology:

> No longer are ethnomusicologists content to record music in the field—to collect data for later analysis in the laboratory. The shift in interest away from music as an object toward music as culture and then as cultural practice has renewed emphasis on 'reflexive, nonobjectivist scholarship.' Reflexive ethnography responds to two related aspects of ethnomusicological heritage. First, it works to redress the insufficiencies of colonial ethnography that positions the ethnographer outside the study community in an Archimedian vantage point from which to view and represent the Other, resulting in what Gourlay called 'the missing ethnomusicologist' […]. Second, reflexive ethnography rejects the modern-era science paradigm that conceives of human culture as wholly objectively observable […], ethnographers attempt reflexively to understand their positions in the cultures being studied and to represent these positions in ethnographies, including their epistemological stances, their relations to the cultural practices and individuals studied, and their relationships to their own cultural practices. Reflexive ethnography is keenly aware of experience and of the personal context of experience.[272]

[271] cf. Merriam (1964); Hood (1971); Philip V. Bohlman, "Ethnomusicology," *Grove Music Online*. *Oxford Music Online*, accessed 20 Apr. 2011.
<http://www.oxfordmusiconline.com/subscriber/article/grove/music/52178>.
[272] Cooley and Barz, "Casting Shadows," 19-20.

These positions are reactionary, in the true sense of the word. Ethnomusicologists have sought to rectify what they perceive to be a history fraught with missteps perpetrated in the name of science. But in doing so they indiscriminately ignore a legacy of scholarship that has in fact supported their cause. Scholars like Alexander Ellis, Von Hornbostel, Kolinski, Blacking, and Rycroft laid a framework for understanding the complexity of many non-Western traditions and for displacing Western art music from its position as a normative standard. Without their detailed analytical studies it will be very difficult indeed to sketch the nature of human musicking in all its diversity.

Leonard Meyer understood this resistance to scientific method as the product of the opposition of universalism to relativism in ethnomusicology.[273] More recently, Albrecht Schneider has described it as "an opposition of experimental empiricism vs. hermeneutic interpretation." [274] But apart from these ideological differences there are also methodological barriers responsible for this divide. For instance, Bruno Nettl claims that the apparent neglect of comparative musicology in the second and third generations of ethnomusicologists may very well be the product of "our inability to find satisfactory ways of comparing."[275] The most wide-ranging approach to comparative musicology of this generation is arguably Alan Lomax's Cantometrics project. Despite considerable emphasis on scientific methods, Lomax's project involved a more or less arbitrary system

---

[273] Leonard Meyer, "Universalism and Relativism in the Study of Ethnic Music," *Ethnomusicology* 4, no. 2, (1960): 49-54.
[274] Albrecht Schneider, "Comparative and Systematic Musicology in Relation to Ethnomusicology: A Historical and Methodological Survey," Ethnomusicology 50, no. 2 (2006): 236.
[275] Nettl, "Comparative Study and Comparative Musicology," 271.

of classification that was modeled on a flawed statistical design.[276] The grandiose claims that Lomax made about culture-areas and the essential features of various regions did not sit well with the postcolonial narratives then in circulation. Some have resurrected Lomax's work in order to get at similar issues in the past decade or so (Grauer 2006).[277] But as we return to questions of universals and evolution we need to interrogate the assumptions of these past paradigms and seek out new methods that speak to new research questions.

This study moves in that direction by once again integrating fieldwork with analysis. But it is fieldwork of a very different kind than that advocated by the post-Geertzian generation of ethnographers. Establishing a more rigorous method for comparative analysis is certainly one goal. Another is to illuminate those strands of culture that are a context for vocal communication and which must inform any rigorous analysis of human communication. To understand a particular tonal grammar, for instance, we need to distinguish between its constituent elements on a scale of determination from absolute to fluid, including the distinctive interactions of cognitive capacities, linguistic propensities (an extension of 1), and the learned conventions or formulae of a particular community. The role of learning is, in each case, slippery. For is language not learned? Are not the gestures of vocalization the product of practice and mimesis? How concrete is the mind, how extended? A biocultural approach collapses the opposition biology vs. culture,

[276] Gage Averill, "Introduction," in *Alan Lomax: Selected Writings 1934-1997*, ed. Ronald D. Cohen (New York: Routledge, 2003).
[277] Victor Grauer, "Echoes of our forgotten ancestors," *Worlds of Music* 48, no. 2 (2006): 101-134.

innate vs. learned; in place of these oppositions biocultural musicology emphasizes embodied processes that are interactive (see Chapter 1).

Nevertheless, we may observe that music cognitivists have tended to emphasize biological explanations over the 'cultural relativism' of ethnomusicologists. But Gary Tomlinson reminds us that, "culture itself is not a cause of human actions, only a context of which they form a part, in which they take on significance […]. [I]n order to understand the actions of people of other cultures […] we must in some way attempt to comprehend, to construct for ourselves, their context."[278] This is the purpose of the ethnographic enterprise: to observe, to participate, and to reach into such horizons without assuming that they are the whole story. Fieldwork should inform but not determine the analysis.

To summarize: the biocultural perspective developed in this dissertation breaks new ground by integrating field methods from ethnomusicology with methods of analysis designed for comparative study and close analysis. Fieldwork was a necessity for this project because the goal was to establish the nature and function of pitch patterning in natural settings. That is, the study was designed around communicative acts outside of the laboratory and as contrast and complement to contemporary theorizing based on Western repertories and music notation. The idea was to study the basic features of pitch patterning in Zulu song using recordings of everyday music rather than commercial or art

---

[278] Gary Tomlinson, "The Web of Culture: A Context for Musicology," Essays for Joseph Kerman, *19th-Century Music* 7, no. 3 (1984): 350-362.

music. This was to counter a tendency in cognitive music theory that has taken 'competent listeners of Western art music' as normative of music making.

The purpose of the fieldwork was to obtain a substantial body of recorded data for analysis and to make sense of the everyday contexts of music making. It was important to make new recordings because so few archival recordings are available, and most are unsuitable for analysis using spectrogram techniques. Using archival recordings also presents other challenges: since it is impossible to determine how these recordings were engineered and mixed, pitch patterning may have been altered in significant ways. Where instruments and ensembles are included it proves impossible to separate out the melodic line of the singer from that of the orchestra. For these reasons I chose to rely primarily on recordings that I could engineer using microphones with a flat frequency response and placing the microphones in a systematic attempt to ensure isolation between voice parts. This is what facilitated the comparative approach to analysis. In Part 2 of this chapter I describe the background for the case study and report on the gathering of data.

PART 2: CASE STUDY

Zulu History and culture

Bantu-speaking peoples have occupied the coastal plain of southern Africa since at least the fifth century CE.[279] At this time they shared the land with hunter-gatherer societies of

---

[279] Gavin Whitelaw, "A Brief Archaeology of Precolonial Farming in KwaZulu-Natal," in *Zulu Identities: Being Zulu, Past and Present*, ed. Benedict Carton, John Laband and Jabulani Sithole (Pietermaritzburg: University of KwaZulu-Natal Press, 2008): 47-61.

Khoisan who had already roamed these parts for millennia. Some of the Bantu adopted distinctive features of Khoisan languages and culture. Those that borrowed the distinctive click-consonants were eventually to form a distinct subset of the Bantu now classified as Nguni.[280] Over the next millennium several regional dialects solidified into four related languages: isiXhosa to the south, isiZulu to the east, siSwati and Ndebele further north. European colonization disrupted their settlements in the fifteenth century, when Portuguese explorers rounded the Cape and initiated Indian Ocean trade with the East. Their way stations and trade routes were the precursor to more expansive colonization. Beginning in the seventeenth century the Dutch and then British colonized present-day South Africa, gradually extending the Cape frontier and later gaining a foothold at Port Natal. They not only traded with Nguni-speakers but also competed with them for land and resources, encountering fierce resistance along the way. In the course of a long series of bloody conflicts the Khoisan were hunted to extinction on the coastal belt while survivors sought refuge in the highlands and the arid expanses of the Kalahari Desert. There were also campaigns against the Nguni. The Xhosa tribes were routed by British

[280] *Nguni* is originally a linguistic term, and historians John Wright and Carolyn Hamilton caution against using it as an ethnic designation: "Continued use of this term as an ethnic designation helps to conceal the conclusion to which recent archaeological research, as well as research on recorded oral traditions of Zululand-Natal points – that the historically known African societies of the region did not 'migrate' into it in fixed ethnic units, but emerged locally from long-established ancestral communities of diverse origins and heterogeneous cultures and languages" (2008, 56). Tim Maggs argues that, "After about AD 1500 the evidence clearly indicates that the Iron Age people of the Natal region were directly ancestral, culturally, linguistically, and physically, to today's black population. There is nothing to suggest that the cultural traits that distinguish Nguni-speaking groups from others within the Late Iron Age originated or developed outside the historic Nguni-speaking regions" (Maggs 1989, 37). John Wright and Carolyn Hamilton, "Traditions and transformations: The Phongolo-Mzimkhulu region in the late eighteenth and early nineteenth centuries," in *Zulu Identities: Being Zulu, Past and Present*, ed. Benedict Carton, John Laband and Jabulani Sithole (Pietermaritzburg: University of KwaZulu-Natal Press, 2008): 69-82; Tim Maggs, "The Iron Age Farming Communities," in *Natal and Zululand From Earliest Times to 1910: A New History*, ed. Andrew Duminy and Bill Guest (Pietermaritzburg: University of Natal Press/Shuter and Shooter): 28-48.

battalions while the siSwati entrenched themselves as a small polity inland of Mozambique. It was the Zulus who bore the brunt of the military incursions in the late nineteenth century and lost their dominions to the British and Boers in the process.

The Zulu kingdom was established in the 1820s. The ‘Zulu nation’ as an imagined community is the product of a complex genesis. Originally a minor tribe among several more powerful states on the eastern seaboard of southern Africa, the inspirational leadership and military brilliance of Shaka Zulu quickly expanded their power base resulting in a major shift in the economy and social structure of the region. A system of autonomous chiefdoms was replaced by a centralized power structure. The *amakhosi* (chiefs) were expected to pay tribute to the king and to support him in times of conflict and necessity. Shaka’s successors Dingane and Mpande continued his political policies and established a Zulu aristocracy. [281] Careful policies of intermarriage and the training of bodies of young men (*amabutho*) with strict discipline and enculturation into Zulu lifeways facilitated a stable state system. This enabled a formidable response to European colonists. By mid-century the British had established the colony of Natal to the south and the Boer trekkers the Transvaal Republic to the north. Both were hungry for land and labor. This culminated in the Anglo-Zulu war of 1879. The British invaded Zululand and deposed the king, sending him into exile. There followed a period of internal conflict that splintered the kingdom and brought it under the full control of the British. They installed magistrates and imposed a hut tax forcing a system of rural-urban migration that has

[281] Wright and Hamilton, “Traditions and transformations.”

continued to this day. When the Union of South Africa was proclaimed in 1910 the Zulu king Cetshwayo had been exiled as a traitor to the Crown; he died in the eastern Transvaal in 1913. The segregationist policies of the Union (1910-1961) further eroded the already diminished holdings of Zulu-speakers in the region and undermined the authority of the king. By this time all isiZulu-speakers in both Natal and Zululand were understood to be Zulus with a single king. During the apartheid years (1948-1994) the Zulu paramount and the *amakhosi* regained limited powers of government. In 1971 KwaZulu was proclaimed a homeland and the architects of apartheid—by extending colonial policies of segregation, cultural essentialism, and ethnic coherence—were expedient in supporting the notion of a Zulu nation under a single king because it furthered their apartheid doctrine.

The post-apartheid government continues to recognize and support the role of 'traditional' leaders in KwaZulu-Natal and has upheld the status of the Zulu monarchy. A sense of pride in being Zulu was propounded by the KwaZulu government under the Inkhatha movement, and to some extent it rejuvenated disappearing traditional practices. Nevertheless, by the 1980s, "[t]he forces of Westernization, industrialization and urbanization, together with the South African government policy of apartheid, […] led to the steady disruption of Zulu traditional social organization."[282] The designation 'Zulu' today describes an ethnic identity that coalesced in the nineteenth century. Not all first-language isiZulu speakers identify as Zulus, but there are nevertheless similarities in

[282] Rosemary M.F. Joseph, "Zulu Women's Bow Songs: Ruminations on Love," *Bulletin of the School of Oriental and African Studies, University of London* 50, No. 1 (1987): 98.

social organization, belief, dress, and custom across a range of groups living in present-day KwaZulu-Natal.[283]

Zulu social organization determines gender roles. Society is patriarchal and polygamous, with most men marrying multiple wives in traditional law. Men hold positions of power and authority in rural communities from the ranks of indunas to chiefs. Gender relations structure social interaction among rural dwellers and determine the sorts of musical performances women, men, and children engage in and have access to. Most genres of music are organized by gender. For instance, men tend to sing isicathamiya and maskandi whereas women more often perform bow music and wedding songs (*isigekle*). There are both cultural and historical reasons for this segregation. For instance, adolescents are socialized in peer groups who must experience important rites of passage together and who continue to act together through life. The experiences of labor migrancy have further emphasized these gender dynamics. On the other hand, *isicathamiya* is an all-male *a cappella* vocal style made popular in hostels and in cities where labor migrancy and segregationist architecture separated men and women. This is evidence of the lengths that the apartheid governments went to in order to facilitate policies grounded both in segregation and in their essentialist vision for 'traditional' cultures. Nevertheless, contemporary rural Zulu societies continue to encourage a division of labor between men and women and enforce strict codes of interaction (*hlonipha*) based on marriage practices

[283] For detailed discussion of the politics of identity see: John Wright, "Reflections on the Politics of Being 'Zulu,'" in *Zulu Identities: Being Zulu, Past and Present*, ed. Benedict Carton, John Laband and Jabulani Sithole (Pietermartizburg: University of KwaZulu-Natal Press, 2008).

and taboos. These relations inscribe themselves in musical practices and are evident in the public performance of gender, as Louise Meintjes has so deftly demonstrated.[284]

Fieldwork in KwaZulu-Natal

Conducting fieldwork in rural areas of KwaZulu-Natal posed many challenges, but the most basic one was finding participants. I had to improvise a variety of methods to find out who was interested, available, and willing; there were no volunteers. In some instances I relied on personal contacts (including family, friends, and acquaintances) to find participants, but the bulk of the work was done by word of mouth and investigations through elected and tribal officials. This meant negotiating the province's two overlapping modes of governance: the system of municipalities with democratically elected officials and the semi-autonomous system of traditional leaders that functions under the department of Cooperative Governance and Traditional Affairs (COGTA). The latter expands on the system of chiefs described above. It is based on the breakdown of regions into tribal areas governed by kings or chiefs. This was a system imposed under British colonial rule in the nineteenth century but which is nominally based on existing chiefdoms. With land redistribution under apartheid and the resettlement of large numbers of Zulu speakers in areas far from their ancestral homes, the division of land is no neutral issue. It has caused a great deal of social and spiritual tension for the Zulu people and has resulted in faction fighting between tribal units. In the Msinga area a

[284] See: Louise Meintjes, "Shoot the Sargeant, Shatter the Mountain: The Production of Masculinity in Zulu Ngoma Song and Dance in Post-Apartheid South Africa," *Ethnomusicology Forum* 13, no. 2 (2004): 173-201.

pitched battle was fought as late as 1952 between the Mchunu and Thembu tribes. The loss of land and livelihoods and the imposition of hut taxes in the late nineteenth century put a great deal of pressure on rural Zulus to find work in urban areas. This was supported by the government, which encouraged rural-urban migration for cheap labor in the mines of the Witwatersrand, Kimberley, and elsewhere, and in industrial complexes at major cities and ports like Durban and Richard's Bay. The fragmentation of Zulu lifeways and economy was already in place by the mid-twentieth century. The urban-rural dichotomy entrenched as a form of double-consciousness is reality for many if not the majority of South Africans. Work in the cities, home in the country. Labor migrancy gave rise to some of the most popular music styles still practiced today, including isicathamiya, maskandi, and various hybrid forms of afrogospel. This mix of idioms and cross-pollination of rural and urban lifestyles features strongly in the music I recorded. It is obviously meaningless to assume a 'pure' Zulu music. What needs to be recognized is simply that rural musics are not entirely rural but the product of a constant flow of ideas and influences disseminated through the media and migrant labor.

'Traditional' life (and music) coheres through cultural and political systems that emerge from this mix. The 'homeland' policy of apartheid created 'Bantustans' that were operated through a system of governance reliant on 'traditional' leaders. A single *Inkosi* or Chief performed judicial functions and received taxes from their constituencies or *izigodi*. Each *isigodi* was in turn run by an *induna* or headman. He was chosen at the discretion of the chief and the tribal elders. The system continues in the post-apartheid period, with traditional leaders receiving salaried compensation for their duties, (minimal) financial support for the development and administration of their constituencies, and representation in government. Traditional leaders remain power brokers who must be consulted in events of ritual significance. For instance, for any *memulo* (girls' coming of age ceremony) or wedding, a family is supposed to get permission from the chief and pay a levee. Usually the chief will work through *indunas* who represent him at events or public gatherings. These *indunas* also serve as custodians of the peace. They are appointed by the tribal elders on an honorary basis, but their duties may be extensive in many rural areas. I consulted several *indunas* to obtain permissions for my studies and was often expected to make offerings of good will. On other occasions I would meet with the secretary to the *inkosi*. Usually I would leave a letter of intent for consideration by the *inkosi* and *indunas* (elders). Once this had been discussed and agreed upon I was given advice on how to proceed. In some cases this proved crucial to my organizing recording opportunities. For instance, in the Msinga district I was given access to the list of *memulos* and *umshados* (weddings) scheduled for the Easter holidays by the Mchunu and Thembu tribes.[285] In other places the situation was more difficult. For instance, I was once instructed to provide payment of a cow for an audience with an *inkosi* in northern KwaZulu-Natal!

The other face of rural life is the domain of democratically elected officials and municipalities. Local government counselors proved helpful in many instances. In fact

[285] All *memulos* have to be sanctioned by the *inkosi* who must be paid a levy in advance.

they invariably sought out and guarded the rights of their communities. I had to explain my case to these officials, sometimes in the middle of a recording session, before I could proceed. Municipalities have become increasingly powerful because they consist of elected party-political representatives. Traditional leaders belong to a system of patronage and authority that is losing ground, especially among the younger generation. This division of power has resulted in tensions between the traditional and political substructures in rural areas, and in some cases it complicated my fieldwork. Determining the correct protocol became more complicated as I sought permissions from authorities. In the end I resolved this according to the sort of fieldwork being done. When I recorded traditional events that were under the authority of the chief I had to begin with the *inkos*i or *induna* for the area. But in areas like Ndumo, where traditional functions were rare and the adult population scarce, I focused on schools instead. This meant work with educators and counselors rather than chiefs. Working through these various traditional and governmental bodies proved useful, but it had to be complemented by a more time-intensive method that essentially involved working by word of mouth. This is demonstrated by my experiences in northern Zululand in early 2012.

Setting out

The unpredictable nature of fieldwork is demonstrated by my second stint in late January-February of 2012. I traveled to Ndumo on an exploratory trip in mid-summer. It was extremely hot, humid and mosquito-drenched work. My guide and interpreter at Ndumo was an old family friend, Thulani 'Roger' Gumede. He had arranged a few meetings for me in advance, and several of these proved fruitful, but the best practice was to drive from homestead to homestead, following leads as best we could. The musicians and dancers in the area were mostly well known, but finding them often proved to be a major challenge. The very first maskandi artist that we met was a football friend of Roger's, Mduduzi Ngomezulu. He is a brilliant guitarist and lyricist who performed an impromptu concert for us at his *umuzi* (homestead) near Makanies. He promised to record scores of songs for me and was very excited about the prospect of a 'demo' CD. He had been saving to record a demo at a local studio but had not yet come up with the R400 needed.[286] My arrival was thus a welcome opportunity. In fact, there were many, many other groups who responded in the same way. I became their free entry into the music industry, or so they hoped. By the end of my first trip Mdu and I had reached an agreement that I would record him with his group when I returned in March or April. But each time I came back—four times that year—he was inevitably away working as a laborer at sugar cane plantations near Empangeni (a city several hours drive from Ndumo). And he was not the only one. Most men and women who were healthy enough left in search of work on farms, in cities, or in industrial employ. This proved to be a major problem for my fieldwork. The only way to get around it was to record over holidays and to work with school children and the elderly. But no appointment was guaranteed and my schedule was stretched to the limit. During this first trip, for instance, I also met with Sibusiso Khumalo. He taught *indlamu* dance at Skhemelele and was coach of the provincial champions. Yet despite three further trips to the area and many

[286] R400 is equal to roughly $40 as of June 2013.

meetings with him, I was unable to record him or his group even though I attended practice sessions and one performance at a competition.

The challenges at Ndumo were indicative of many more to come. Transport was a problem because we had to locate people in areas with few roads and no signage. The car I initially borrowed was damaged in an accident at Msinga and took two long months to repair. I was forced to hire a car that was not designed for off-road driving and had to navigate treacherous stretches of sand and gravel on instinct. I eventually hired a driver, but he was unfairly arrested for possession of an assault rifle and several weeks in jail. Further delays ensued. Fieldwork, in short, was not simply a case of field recording.

Sites and recordings

Initially I planned to set up six main sites in KwaZulu-Natal that ranged from North to South and East to West to ensure a wide sample of Zulu songs and to enable a comparative perspective. The site list included the districts of Amazizi, Msinga, Ndumo, Nongoma, Umkomaas, and Himeville. I visited all of these but in the end made just a single recording at Umkomaas, and just one event at Nongoma. Fortunately I was able to set up recordings at Dakeni and Eshowe in southern Zululand that substantially improved the sample. I also traveled to Vryheid in central Zululand to study bow music with Brother Clement Sithole. He taught me basic techniques for building and performing *umakhweyana* and *ughubu* bows. I was to return to record him and his students in September but fell ill with fever on arrival in Vryheid and had to cancel. In the end the total number of sites expanded to include ten districts: Amazizi, Mweni, Msinga, Weenen, Himeville, Dakeni, Eshowe, Ndumo, Nongoma and Tembe. The first recordings were made at a *memulo* at Ncunjane in the Msinga district on 28 and 29 December 2011 and the last was made at Mashunka on 2 September 2013. A complete table of audio recordings is listed in **Appendix A** and audio-visual recordings in **Appendix B**. Many other archival and commercial recordings of Zulu popular and neo-traditional music are available. The most important for this thesis are recordings of maskandi and isicathamiya, some of which are discussed in more detail in Chapter 4.

I very quickly had to establish a flexible field recording technique. At most traditional events I had to film and record audio simultaneously. Monitoring recording and noise levels, microphone placement, and the events unfolding was a constant juggling act. The decision to do both came of necessity, because my attendance at events usually entailed some form of transaction. I would be allowed to make audio recordings if I filmed the event and edited and distributed DVD copies to the family afterward. In some cases I was both videographer and photographer and sometimes found myself filming others filming.[287] This practice led to some animosity on the part of rival videographers, who felt I was undercutting them (I was of course performing this duty free of charge). There were nevertheless some benefits to this filming. Not only did I gain an extra high

[287] This was not always a positive experience. At one *memulo* in the Drakensberg I was forced to stop filming and told to leave under very difficult circumstances (this after having waited months for the event and having traveled more than five hours by car).

definition audio recording, but I also documented the close relationship of song, dance, and gesture.

Less than a third of the days spent in the field actually involved recording, and most of my time off from fieldwork involved editing, rendering, and systematizing data. Because most traditional events take place over multiple days I obtained a massive amount of video and audio data that had to be downloaded, edited, and rendered even before analysis could take place. For the sound recordings I had to reduce noise levels, cut out unwanted sounds, mix tracks and check balance levels. The video editing was complicated by software difficulties that made it difficult to render PAL-format DVDs. The rendering of each video took on average two to three hours. Another complication was that most rural homes have only very basic playback facilities. On follow-up trips I had to make deliveries and check that the DVDs and CDs actually worked on their equipment. Sometimes I had to return two or even three times, and on one occasion I eventually bought a new DVD player for a family in Ncunjane so that they could watch their wedding video!

The audio recordings for this project were of two main types based on location: those that took place (1) outdoors and (2) indoors. The former were usually also filmed. Recording in unpredictable and sometimes harsh environments was challenging in several respects. For instance, it was nearly impossible to control for environmental noise, rain, wind, livestock, and birds, and the excitement of the sometimes quite large crowds in attendance. At Zulu *memulos* and weddings I would have to move around a lot during the proceedings. I filmed using a handheld Canon high definition video camera and recorded with a Zoom H4N digital recorder with several microphones. For outdoor events I relied on a shotgun microphone with a shock mount and wind shield. Sometimes there would be singing emanating from two or three groups simultaneously and I would have to choose between them. There was no chance of re-recording these events later or of capturing individual voices, and so important decisions had to be made on the spur of the moment. More controlled recordings were only possible on request and had to be made indoors.

Recording indoors was no less complicated. Many venues were not well suited to recording. For instance, the government-built community halls and schools in rural areas seldom have electricity, lighting, or windows intact, and the corrugated iron roofing and concrete flooring echoes with no ceilings. The best results came from recordings made in *rondavels* or traditional circular huts with polished dung floors and thatch roofs. At schools and community halls my recording sessions were usually of great curiosity to the community and passers by. An audience would gradually build up who, despite their best efforts, could not contain their excitement and curiosity. The results were, in retrospect, sometimes comical. Invariably children would climb up the walls and hang through open windows, there would be babies crying on mothers' backs, excited ululations from elderly women (as is customary), and a host of other 'interferences'. Then there was the question of microphone technique and placement. It often required several takes before singers and instrumentalists were comfortable with the various sorts of microphones I used. At least a third of all recordings were deleted. All of this led to a great deal of frustration on my part, and I had to drastically rethink how the materials I collected would or could be used to address my research questions.

Reciprocity

While in the field I had to practice diplomacy in conflict situations. At a remote homestead on the border with Mozambique I was intimidated by a group of twelve men carrying traditional weapons—including spears, shields and clubs—who insisted on a large down payment for their services (which at that point remained unknown). They performed a war dance to demonstrate their skill and perhaps also their not insignificant power. Despite Thulani's best attempts to placate them, and to explain the purpose of my work, this group insisted that I pay them R10,000 for recording just four of their songs. And they had some justification for this request. Fifteen years ago they had performed at a government event for which they were paid a similar sum. Their reputation was considerable, I was told, and they performed with famous maskandi. Unfortunately I had no way of corroborating these claims and certainly did not have the resources to accommodate them. Negotiations of this sort were common and were the result both of misunderstanding and of understandable opportunism. It was very difficult to explain academic goals to musicians for whom the concept of academic research meant almost nothing; and for whom the outcomes had no value. In many cases, however, these negotiations were complicated by precedents that had been set by researchers and organizations who had told different stories. For instance, at Nkaseni a political party paid a primary school group R7,000 for a dance performance. For me to conduct research on these terms I would have needed a budget of tens of thousands of dollars, more than was available to me as a graduate student.

All of these examples point to the need for reciprocity in fieldwork. I felt the need to find material ways of contributing to communities and so I came up with various strategies. Following consultation with NGOs working in rural KwaZulu-Natal, I decided to pay all groups a basic hourly fee, and I provided many with food and drink as well. For traditional ceremonies I provided multiple edited DVD recordings in place of a videographer. This meant follow-up visits, of course, and additional expense. I traveled more than 25,000 kilometers during the fieldwork phase and spent many thousands of dollars on petrol and accommodation. I also employed interpreters and guides at Ndumo, Msinga, Mweni, Amazizi, Dakeni and Himeville. Their mediation was crucial in many situations of conflict and opened access to tightly knit communities, many of whose members were suspicious of me. Traveling with a companion was also a safety precaution. But these measures all had a considerable financial burden and there were cases where I could not conduct fieldwork for months while I waited for funding to come through.

A further challenge was dealing with other field researchers, some of whom worked prior to my arrival and others who worked simultaneously. In one area I had to contend with expectations for remuneration and promotion that were well beyond my budget. In a second area a sociologist had not paid cash but had provided food and drink, including meat. My consumable offerings were less substantial, and the remuneration was not immediate.[288] A common demand was for alcohol. Sometimes this entailed commercially produced vodka (1818 or Smirnoff, beer, and the like) but other times it had to be brewed to invoke the spirits. When *amahubo* (anthems) are sung the spirits are present and this usually means the brewing of traditional beer. I was uncomfortable paying in liquor and so lost out on many opportunities to record men singing traditional songs. Another challenge was introduced when a researcher in my region arrived and paid interviewees substantially more than I could afford. Despite our vastly different research areas I could not gain entry without paying an equivalent rate. This says a lot about the shadows one casts as a researcher and the impact of research on the community as a whole. With the ever-extending reach of the academy it is becoming increasingly necessary to establish accepted standards for scholars across disciplines who engage in community work.

Challenges

Traveling in unknown and remote territory posed many challenges. In South Africa, with its complex history of violence and unrest, mistrust and neglect, my meetings were often very complex cultural negotiations. For the most part I found people to be interested, polite and friendly. Where there was intimidation and harassment—sometimes I would receive late-night or early-morning phone calls and abusive text messages—it was usually the result of anxiety and misunderstanding. The success of all meetings depended

[288] I made bank deposits or transferred funds to cellphone bank accounts via local stores afterwards so as to protect myself from theft.

on our reaching a basis for trust and I did not make a single recording without obtaining informed consent.

Racial prejudice proved a major encumbrance and was the single greatest impediment to my research. It was not simply that I was 'white' working on 'black' music. More specifically, to nearly all those with whom I worked, I was an *umlungu*. The term is used to refer to 'whites' in derogatory fashion. It is freighted with historical injustices reaching far deeper than the radical inequality of racial prejudice and economics. Wars fought, land dispossessed, livelihoods destroyed and uprooted are the realities of South African history. Stories of exploitation, brutality, and hatred have scarred communities living in abject poverty. It was difficult for me to counter these narratives in communities where this suffering ran deep. I spent many hours listening to harrowing stories and *songs* about apartheid abuses, cruel farmers, shocking wages, and a slave-like system of exploitation.[289] Sometimes songs were coded with messages of discontent and anxiety that I only picked up on much later when editing and analyzing the material. Through continued contact and communication many of these frustrations were overcome.

Throughout my fieldwork I worked on a basic principle of reciprocity: that community members benefit in tangible ways that suit and accommodate their needs, and that they are fully informed about the goals and ends of my research. I devised a form in

[289] At Nkaseni in central KwaZulu-Natal I was told of how people living on white-owned land had to work six months of the year without pay, just to be allowed to keep their homes and graze their cattle.

consultation with Rauri Alcock and the NGO C.A.P. that used elements from the Smithsonian Folklife document on informed consent for fieldworkers. I set up parameters for the negotiation of such agreements, but my printed documents were often unintelligible to musicians. The documents were translated into isiZulu and were read to all participants before recording took place. In retrospect it seems to me that my intentions were not judged on the basis of these documents alone, but rather on my delivering on my promises.

There were two starkly contrasting positions that emerged from these discussions: the first was the suspicion that I would exploit musicians and run away and make millions on the world market. The second was the hope that I would go and make millions for the artists—i.e. that I would produce and market recordings to make money for the artists concerned. Those groups who were more serious about music-making as a way out of their current jobs or lack of employment were desperate that I promote them. The fact that I came from ‘overseas’ was no doubt important in their estimation of my work, but I could not guarantee any promotion of this kind. In sum, my efforts to achieve reciprocity were time-consuming and costly, but they were an important component of this project and so are recorded here.

PART 3: MUSIC COGNITION IN THE FIELD

Figure 2: *Memulo* at the Mbhele *umuzi* in Dakeni, KwaZulu-Natal (10 July 2012)

"By creating a reflexive image of ourselves as ethnographers and the nature of our 'being-in-the-world', we believe we stand to achieve better intercultural understanding as we begin to recognize our own shadows among those we strive to understand".[290] Being

[290] Barz and Cooley, "Casting Shadows," 4. Over the past several decades increasing specialization in the music disciplines has resulted in an unfortunate disconnect with non-Western cultures being studied primarily from an ethnographic rather than a musicological perspective. This trend is evident in literature on South African music in particular. Many scholars avoided perceived 'ethnocentric' topics of investigation as a response to the state's insistence on segregation and separate development. As a result many scholars turned away from studies of 'traditional' music and toward broader sociological perspectives. Studies of popular music and jazz focused on topics of politics and social change. The focus of ethnomusicologists in the United States has increasingly been on music *in* culture or music *as* culture.

in the world is seldom a preoccupation of musicologists, nor music cognitivists. Pouring over scores, transcribing recordings, scrutinizing the archive, laboratory work: these are solitary pursuits. But if scrutinizing one's object of inquiry requires fieldwork then one encounters very different challenges. Cognitive musicologists who conduct fieldwork cannot, and should not, ignore the contexts and consequences of their work in communities. Fieldwork is an unpredictable enterprise and its scholarly and personal outcomes are not guaranteed. Field workers must plan for the potential impact of their research projects in social and material terms. Where fieldwork is collaborative and takes place with communities for whom academic research is apparently irrelevant, achieving reciprocity means making substantial material investments. In vulnerable communities, where day-to-day realities present critical challenges to subsistence it is unreasonable and indeed unethical to expect communities to commit resources and effort to activities that do not accrue a tangible benefit. Simply documenting cultural practices is no longer enough to guarantee access, or to meet the need for reciprocity. The appeal to posterity—which has for so long been the standard refrain of anthropologists and historians—is simply untenable in present socio-economic circumstances. In South Africa in particular, the weight of history crushes naïve assumptions about the intrinsic value of research and the ends to which it is put. The ethical conduct of researchers is closely monitored and regulated within academia, and by communities. For the latter, trust must be gained over time by establishing committed relationships. No credible research can be undertaken

---

This dissertation reasserts a musicological perspective by returning to the analytical approaches of an earlier generation, including J.H. Kwabena Nketia, David Rycroft, John Blacking. At the same time, I recognize the need for scholars in cognitive musicology to take account of the social consequences of their work. I reflect on my own fieldwork experiences and argue for a socially responsible musicology.

without informed consent and approval established in advance. In my experience it was necessary to demonstrate in concrete terms how communities would benefit and how their rights would be protected before they would contribute to my work. This involved providing official information on the project as well as signed documents of informed consent that specified the uses to which research data would be put, and how participants would benefit.

Fieldwork is a humanistic process that is not reducible to its objects of inquiry. It is a process of social interactions, of negotiation, of learning and application. It is protracted and its processes and products cannot be accommodated to the norms of laboratory research, simply because its outcomes are largely beyond the fieldworker’s control. There may be an event this week, this month, or this year, or there may be nothing to record over the next two years. It is therefore an unpredictable enterprise that is difficult to plan for. Budgets and timeframes must be flexible if success is to be had.

All of these factors point to the tension between scientific methods of inquiry and the goals of reflexive, reciprocal research. The goals of research in music cognition are very different from those of ethnomusicology, even if both sets of researchers are sensitive to the processes and social consequences of their fieldwork. In ethnomusicology, the many challenges raised in Part 2 of this chapter are an important part of the work ethnographers do, and this emerges in how they think and write about it. This is not the case for music cognitivists who report only on the data they have collected and the results of their experiments and/or surveys. In short, there is a less productive tension between fieldwork and publication for music cognitivists because the *work* of being in the field is not integral to the *goals* of the project; in fact, the practice and process of fieldwork may disappear entirely from the published results. I have tried to retain the traces of this work both in the analyses of the previous chapter, and in this report. One way to get around this is for music cognitivists to become ethnomusicologists, or to contribute to publications on the practice of fieldwork, and the role of social contexts in understanding music perception and cognition. The usage-based approach adopted in this dissertation claims that we need socially embedded studies in order to understand the communicative contexts out of which music is made, communicated, and understood. Rather than minimize sociality we need to understand how it provides the ground from which music emerges.

CHAPTER 6

Toward a usage-based theory of music

This dissertation contests a series of assumptions that have gained currency in cognitive music theory through the dominance of a rationalist mode of inquiry. I have showed how considerable accumulated evidence from psychology, neuroscience, linguistics and music cognition undermines the nativist position, and how a biocultural approach to the study of music and mind—that is, one that takes into account the cognitive mechanisms that structure our engagement with the world but also the rich cultural contexts in which social learning takes place—provides an exciting alternative. This usage-based framework treats song as a gradient, adaptive system that is irreducible in its complexity, and that is the product of biological and cultural interactions through ontogeny.

In Chapter 1 I showed that prosodic features are one important linkage between song and speech. Studies of music and language acquisition both show that pitch patterning is prosodic in its earliest manifestations and that mothers use a form of infant-directed speech that accentuates gradient features of melody to communicate effectively and affectively with their children. Pitch categories are not pre-determined by an innate mechanism but are learned through processes of enculturation and training. These categories are culture-specific, even if the mechanisms that guide their use are not. Studies of evolution and universals have also pointed to pitch prosody (especially intonation) as a potential origin for our abilities for complex, combinatorial thought. The paradigmatic elements of pitch may be of a different order to language even though studies of tone languages show important similarities even in this domain. The underlying mechanisms that structure music perception and cognition seem to be shared by music and language (Patel 2008). Neuroscientific evidence shows considerable overlap in the processing of music and language in the brain. These findings—and the fact that multiple areas of cortex are recruited for music and language processing—undermine the notion of specialized Fodorian modules proposed by nativists and evolutionary psychologists. The nature of our current evidence supports the hypothesis that we have *domain-general* learning mechanisms that are not locally instantiated or specialized. Abilities for intention-reading (theory of mind) and patterning-learning (categorization) enable us to acquire music and language in any culture into which we are socialized without pre-specified knowledge. These *social*-cognitive mechanisms enable us to extract regularities through usage, learning as we go, and to construct the grammatical dimensions of music.

Throughout the dissertation I have been critical of ethnocentrism in music cognition. In Chapters 1 and 2 I argued that most studies in the field today rely almost entirely on models based on the structures of Western art and popular musics of the common practice era. Even more problematic is that the categories of staff notation are used unreflexively as the basis for experimental designs that model music and mind. But the categories embodied in notation are culture-specific and reductive. They ignore elements of gradience and variation that we find in all world cultures and that characterize prosodic phenomena in song and speech. These are the same models used by formalists to model music and mind. They argue that there are universal rules of grammar that structure music and linguistic thought across cultures, but attribute these to a genetic inheritance rather than a social one.

A more thorough critique of rationalist models in cognitive science is developed in Chapter 2. The first two parts of this chapter focus on generative approaches to music theory and linguistics, and how the conceptual framework that they have developed has prematurely restricted our investigations. Rationalist epistemologies make assumptions about the structure and functions of the mind that are the product of computationist thinking. The mechanistic conception of the mind as a digital input-output system has no analogy in biology, and neither should it in studies of communication. There is no evidence that the mind-brain processes mental symbols in a time-independent fashion, or that there is a universal grammar that combines these discrete symbols into meaningful utterances. Nativists tend to deny the flexibility of the human cognitive apparatus and underestimate the role of domain-general capacities in structuring thought and action. We must jettison the preoccupation with digital systems if we are to understand the gradient and prosodic features of human communication. This also means adopting new methods to investigate the properties of music and mind we thought we knew through staff notation. Advanced recording technologies and techniques used in conjunction with spectrographic and pitch tracking software enable a more fluid accounting of musical performances.

Human musicking is processual, non-reductive and complexly ordered in both syntagmatic and paradigmatic domains. Empiricist approaches to music and mind take these factors into account by modeling song as a *complex adaptive system*. This system is adaptive because it is based on past experience, and past and present experience feed into future usage. It is complex because fluctuations in pitch (and other parameters) simultaneously convey a rich diversity of information as part of a multi-dimensional, networked structure. This structure is not decomposable and is best conceived as a 'flat hierarchy' consisting of multiple interacting operations.

The usage-based approach also stresses the need for studies of process and performance. Rationalists have reified a competence vs. performance dichotomy that ignores the diversity of possible performances and the ways in which acquisition is shaped by usage. Formalist approaches take as their starting point the musical surface—i.e. the notated features of the musical score. But this is an impoverished view because it fails to recognize the score as a culture-specific technology for recording Western art music. We cannot assume that the categories it reifies are grounded in cognitive reality. This point is supported by the examples discussed in the case study of Zulu song prosody. We see how the pitch categories enshrined in staff notation fix the pitch categories of Western music performances artificially. These 'discrete' categories cannot account for the gradient pitch categories used so pervasively in Zulu song.

Chapter 3 outlines a theory of melody as prosody. The first part of this chapter reviewed literature on melody. I demonstrate that the usage-based approach has important precedents in the Penn school of music theory, especially the work of Leonard Meyer and Eugene Narmour. Their emphasis on social learning and Gestalt principles, in tandem, provides an important platform for this work. In this section I also introduce several new pitch categories to account for elements of gradience in song melody, including pitch targets, contours, and 'chunks.' These provide an important link to the prosodic features that I introduce the second part of the chapter. In this latter section I draw on a wide range of research in phonetics and phonology to show that pitch patterning in song is a multi-dimensional combinatorial system comprising melodic categories that are gradient rather than discrete, and which are of a kind with prosodic (i.e. linguistic) pitch categories. Focusing on melody as prosody in this way draws attention to the embodied nature of song as a communicative system.

Music theory tends to derive laws of melodic structure from abstract principles—based on an imposed system of grammar—rather than on the physical features of actual performances. But the *direct* evidence of song usage that I have present in Chapter 4 shows that phonetic cues are critical to perceiving and encoding melodic structures. Pitch movements conditioned by vowel height, duration, alignment and scaling are incorporated into this model, and the voice, as an embodied gestural communicative system, is re-introduced as the basis for our investigations. It seems to me that the disconnect that has arisen between music theory and phonetics is a consequence of our relying too heavily on the pitch structures imposed by common practice tonality and its system of notation and analysis. The syntactic relations codified in common practice theories—such as Schenkerism and the generative theory of tonal music—are based on a notion of musical *competence* that *cannot* be used to derive the outputs of an *actual performance*. Phonetics provides not only the instrumental tools for modeling melody, but also the conceptual tools for making sense of its physical properties.

In Chapter 4 I present a case study of Zulu song prosody. The rationale for this particular case study was twofold: first, I aimed to counter the ethnocentrism of current music cognition by conducting a detailed study of a non-Western music, and using phonetic methods that are themselves culture-neutral. Second, I aimed to show how the pitch parameters of a tone language share many prosodic features (including the new pitch categories introduced in Chapter 3) with song. Zulu song is a good case study precisely because of the profound import of linguistic pitch determinants. It also displays elements of gradience in its pitch categories and intonational structure that impel us to find alternative cognitive explanations for melodic structure. The Chapter begins with a review of earlier literature on the pan-stylistic features of Zulu song with a focus on melodic structures. I argue that the bias toward 'harmonic' explanations of melody so common in music cognition—and in Africanist music theory—are largely the product of

Western ethnocentricism rather than hard evidence. That is, to explain the structure of pitch patterning in terms of harmonic systems derived from instrumental music is problematic given: (1) the paucity of instruments, (2) the sheer variation we find in song, and (3) the considerable evidence of *linear* prosodic elements in this structure. Zulu song prosody consists of suprasegmental elements that are both syntagmatic and paradigmatic, and so we should avoid reductive vertical-harmonic analyses.

A case study of ten *memulo* songs provides the empirical basis for many of the arguments developed in Chapters 3 and 4. I show that pitch is not employed in the discrete or contrastive fashion proposed by many music theorists and most (generative) phonologists. There are a range of pitch categories that include level pitch targets—which emphasize the fact that these are abstract phonological categories rather than the level tones we sometimes find in a pitch surface—but also glides and contours. Melody has a multi-dimensional pitch structure that is inclusive of prosodic features of speech tone and intonation, as well as music-specific paradigmatic elements (a differentiated pitch space with multiple pitch targets). I use several graphing techniques developed in Praat to show the close relationship of song and speech contours, and I present evidence that shows how these were mostly the product of speech-tone determinants. In isiZulu these speech tone features include depressor consonants, contrastive use of High and Low tones, glides, and down-step. The most significant feature of intonational structure, on the other hand, is downdrift. This operates within and across phrases and cycles and results in a gradually constricted pitch-span. This is an important finding because it shows that our perception of melodic structure is not bound to discretely organized pitches, but that we are able to accommodate changes in pitch-span and identity that fluctuate. In other words, we rely on multiple pitch cues to construct an accurate melodic percept.

Chapter 5 was designed as the complement to the case study of Zulu song prosody. All of the recordings I used for my analyses were recorded during a period of fieldwork in KwaZulu-Natal, South Africa, that began in December 2011 and that ended in September 2013, with most of the work completed in the first year or so. This Chapter provided background on Zulu culture and history that also set up my discussion of fieldwork as a social process. Part of this chapter was dedicated to a discussion of cognitive science approaches in ethnomusicology. These have largely been phased out since the demise of comparative (and to some extent, systematic) musicology several decades ago. This dissertation advocates a return to analytical studies of world music that are grounded in the interdisciplinary thinking of cognitive science (rather than, say, the pseudo-scientific methods of Lomaxian Cantometrics), but I recognize that this poses considerable challenges for theorists whose goal is not ethnographic or 'thick description,' but rather the collection of 'data.' However, if we shift our perspective and adopt a usage-based approach we see that investigations into the social conditions out of which music emerges are critical to understanding music and mind, and can only be accomplished through committed fieldwork. This commitment should involve not only analytical rigor, but also an ethics of reciprocity. The social dimensions of fieldwork shape the recorded outputs of the fieldwork experience in important ways, and they need to be recognized and credited.

Although each of the chapters in this dissertation advances an empiricist approach to the study of pitch patterning in song, I do not of course discount the role of rationalism in the study of music and mind. But if we are to make progress that is not merely conjectural—as much generative theory has been for several decades now in its attempt to assemble a universal set of rules and procedures—then we must ground our work in empirical research; we should not, in other words, proceed from conjecture to empiricism in order to fit out data to our theories.

Fieldwork is of course very important to a usage-based theory of music because the focus is on usage and performance. But it also offers an important counter to a discipline that has largely ignored detailed studies of non-Western musics. Music cognition has lagged well behind other disciplines in cognitive science in establishing a cross-cultural theoretical framework. Unlike investigators in psychology, anthropology, and linguistics, music theorists have for too long now retained an ethnocentric focus on Western art music of the common practice era. This focus has been compounded by an allegiance to methods of analysis entirely reliant on staff notation. These are the sorts of basic methodological, theoretical and epistemological challenges facing contemporary cognitive music theory today. The field continues to be dominated by laboratory-based research and funding for experimental science. If we continue on this course—studying small samples of highly skilled student populations at North American and European universities—then we are unlikely to encounter and make sense of the diversity of world music traditions and their systems of tonality. My case study shows that African tone systems possess features that do not fit well with our Western models, a fact that linguists have recognized and investigated for decades already. This reinforces the point that current modes of investigation cannot continue to draw conclusions about universals and evolution, or to theorize the musical mind, without taking into account such non-Western traditions.

***

In many ways this dissertation offers a *preliminary* analysis of a complex set of issues. The case study itself is only a first step in analyzing a much larger data set and in integrating a theory of pitch categories into a theory of prosodic structure. However, the approach that I advocate in Chapter 1—and that has guided this endeavor in all its manifestations—is best understood as the product of a usage-based theory of music. This framework is guided by several principles and convictions that I summarize here.

A usage-based music theory studies the production *and* perception of music so as to avoid listener-centric approaches focused on competence. Elements of production include physical properties that are embodied in performances. These performances should be studied in situ rather than only in laboratory settings so as to take account of the ways in which the grammar of musical structures emerges from particular social conditions. Controlled experiments are important, but we should supplement these with fieldwork-based studies that explain the rich cultural environments in which music is learned and passed on. We cannot afford to restrict analysis to notated knowledge alone. A usage-based theory of music must integrate theoretical research on the cognitive mechanisms that structure perception and cognition with applied research on the *social* dimensions of learning and dissemination. The diversity of music making across cultures must be studied in detail, not simply through selective and predictable sampling of world music traditions, and superficial cross-cultural analysis. We need more detailed data sets and scholarship that is culturally informed and committed. The relationship of music to other behaviors, including language, dance, and religion must also be investigated so that domain-general mechanisms are better understood, and their application to music is clear. Usage-based models must investigate music in everyday life (amateur music-making) rather than elite music making in the academy. If we are studying music as a cognitive capacity then we need to establish how people use it as a daily mode of interaction rather than only as an art form. This also means guarding against ethnocentrism in our methods. New technologies and techniques are available that are less restrictive than staff notation.

All of these principles focus on *usage*: how music is made, performed, and experienced in real-time in everyday lives. To understand the vastness of human music in our diverse behavioral ecologies, and to understand its origins and history, we must expand our view. A usage-based theory enables us to achieve this because it equips us with the skills and biocultural knowledge we need in order to study music as the complex adaptive system that it is.

APPENDIX A: SOUND RECORDINGS

Group/Performer/Event	Genre	Date	No.
1. MSINGA			
A. Ncunjane			
Ncunjane Memulo	Memulo/Amahubo	28 December 2011	22
Ncunjane Wedding	Wedding/Amahubo	29 December 2011	10
B. Mashunka			
Mashunka Memulo	Memulo/Amahubo	5 January 2012	12
Mashunka Memulo	Memulo/Amahubo	6 January 2012	7
Umuzi of N. Dladla	Memulo songs	7 February 2012	15
Umuzi of N. Dladla	Wedding songs	12 April 2012	10
Umuzi of N. Dladla	Isigekle	12 April 2012	6
Khonazugcina	Maskandi	7 February 2012	7
Mdidiyeli	Maskandi	7 February 2012	14
Mvezi	Maskandi	7 February 2012	5
C. Ngongolo			
Mathashi	Wedding songs	10 April 2012	3
Ngongolo Maskandi	Maskandi	10 April 2012	10
Ndlovu Enthathu	Isigekle	10 April 2012	4
Maskandi	Maskandi	13 April 2012	8
2. WEENEN			
A. Thukela Estates			
Imbongi	Izibongo	13 April 2012	2
Lethukuthula Brothers	Isicathamiya	13 April 2012	7
Abafana Bothando	Isicathamiya	14 April 2012	8
Abashana Maskandi	Maskandi	14 April 2012	4
Amaveza Maskandi	Maskandi	14 April 2012	3
Maskandi Amavezokuhle	Maskandi	14 April 2012	3
Noweshe Nyawose	Maskandi	14 April 2012	9
Phendukile	Imfiliji (harmonica)	14 April 2012	4
Kwamlungu Girls	Girls songs	14 April 2012	5
Indhlamu	Indhlamu	14 April 2012	3

B. Nkaseni			
Mbhele Memulo	Memulo/Amahubo	8 April 2012	11
Mbhele Memulo	Memulo/Amahubo	9 April 2012	11
Abacabangi	Maskandi	15 April 2012	12
Sizanani	Isicathamiya	15 April 2012	11
3. AMAZIZI			
A. Cavern			
Cavern Choir (live)	Afrogospel/Wedding	4 January 2012	7
Cavern Choir	Afrogospel/Wedding	11 February 2012	15
Bergville Blue Roses	Isicathamiya	11 February 2012	12
Cavern Choir	Afrogospel/Wedding	21 April 2012	15
Bergville Blue Roses	Isicathamiya	21-22 August 2012	12
B. Amazizi			
Florence's Umuzi	Ingoma/Indlamu	22 April 2012	12
Ingoma	Ingoma	22 April 2012	9
4. MWENI			
Ingoma Ladies I	Ingoma	6 July 2012	21
Ingoma Ladies II	Ingoma	7 July 2012	30
Ingoma Ladies III	Ingoma	8 July 2012	18
Maskandi I	Maskandi	6 July 2012	6
Bergville West	Isicathamiya	7 July 2012	9
Indlamu	Indlamu	8 July 2012	3
Memulo	Memulo songs	7 July 2012	1
Maskandi II	Maskandi	14 July 2012	6
Men's songs	Traditional	12 December 2012	14
5. NDUMO			
A. Ndumo Hill			
Ingadla Ladies	Ingadla	30 April 2012	3
Mqamuli Wezintambo	Maskandi	30 April 2012	9
Shaka Dance	Amahubo/Indlamu	30 April 2012	3
B. Ziphosheni			
Dledleni Gumede	Umakhweyana	1 May 2012	4
Mantombi Khumalo	Isitweletwele	1 May 2012	4

Mbangose Gumede	Umakhweyana/Isitweletwele	1 May 2012	6
Nondlela Mkhize	Mqangasi	1 May 2012	8
Nyanisile Gumede	Umakhweyana/Isitwele./Isizembe	1 May 2012	12
Qomaukufa Ndabeni	Isitweletwele	1 May 2012	4
Zenzile Khumalo	Umakhweyana/Isitweletwele	1 May 2012	13
C. St. Philips Primary School			
Ingoma Girls and Boys	Ingoma/Wedding songs	29 August 2012	3
Amahubo Girls	Amahubo	29 August 2012	3
Soloist	Izibongo	29 August 2012	1
D. Impolimpoli Primary School			
Amahubo	Amahubo	31 August 2012	2
Afrogospel	Afrogospel	31 August 2012	2
Soloist	Izibongo	31 August 2012	1
6. TEMBE			
A. Entokozweni Primary School			
Ingoma Girls and Boys	Ingoma	30 August 2012	3
Amahubo	Amahubo	30 August 2012	1
Soloist	Izibongo	30 August 2012	2
7. ESHOWE			
Eshowe High School			
Indlamu	Indlamu	2 September 2012	2
Eshowe HS Choir	Amahubo/Wedding Songs	2 September 2012	9
Starlight Band	Afrojazz	2 September 2012	4
8. DAKENI			
Memulo	Memulo/Amahubo	10 July 2012	9
9. HIMEVILLE			
A. Himeville Anglican Church			
Angelic Teens	Afrogospel	30 June 2012	13
Udumo Gospel Choir	Wedding Songs	30 June 2012	11
Shosholoza	Isigekle	1 July 2012	10
Siyanqoba	Wedding Songs	1 July 2012	10

APPENDIX B: AUDIO-VISUAL RECORDINGS

Group/Performer/Event	Genre	Date
1. MSINGA		
A. Ncunjane		
Ncunjane Memulo	Memulo/Amahubo	28 December 2011
Ncunjane Wedding	Wedding/Amahubo	29 December 2011
B. Mashunka		
Mashunka Memulo	Memulo/Amahubo	5 January 2012
Mashunka Memulo	Memulo/Amahubo	6 January 2012
Dladla Umuzi	Memulo songs	7 February 2012
Dladla Umuzi	Wedding songs	12 April 2012
Dladla Umuzi	Isigekle	12 April 2012
C. Ngongolo		
Mathashi	Wedding songs	10 April 2012
Ngongolo Maskandi	Maskandi	10 April 2012
Ndlovu Enthathu	Isigekle	10 April 2012
2. WEENEN		
A. Thukela Estates		
Imbongi	Izibongo	13 April 2012
Abafana Bothando	Isicathamiya	14 April 2012
Maskandi Amavezokuhle	Maskandi	14 April 2012
Noweshe Nyawose	Maskandi	14 April 2012
Phendukile	Imfiliji (harmonica)	14 April 2012
Kwamlungu Girls	Girls songs	14 April 2012
Indhlamu	Indhlamu	14 April 2012
B. Nkaseni		
Mbhele Memulo	Memulo/Amahubo	8 April 2012
Mbhele Memulo	Memulo/Amahubo	9 April 2012
Sizanani	Isicathamiya	15 April 2012
3. AMAZIZI		
Florence's Umuzi	Ingoma/Indlamu	22 April 2012
Ingoma Ladies Groups	Ingoma	22 April 2012

4. MWENI		
Ingoma I	Ingoma	6 July 2012
Ingoma II	Ingoma	7 July 2012
Ingoma III	Ingoma	8 July 2012
Bergville West	Isicathamiya	7 July 2012
Indlamu	Indlamu	8 July 2012
Memulo	Memulo songs	7 July 2012
5. NDUMO		
A. Ndumo Hill		
Mafutha Enkukhu	Indlamu	30 April 2012
Ingadla Ladies	Ingadla	30 April 2012
Mqamuli Wezintambo	Maskandi	30 April 2012
Shaka Dance	Amahubo/Indlamu	30 April 2012
B. Ziphosheni		
Dledleni Gumede	Umakhweyana	1 May 2012
Mantombi Khumalo	Isitweletwele	1 May 2012
Mbangose Gumede	Umakhweyana/Isitweletwele	1 May 2012
Nondlela Mkhize	Mqangasi	1 May 2012
Nyanisile Gumede	Umakhweyana/Isitwele./Isize mbe	1 May 2012
Qomaukufa Ndabeni	Isitweletwele	1 May 2012
Zenzile Khumalo	Umakhweyana/Isitweletwele	1 May 2012
C. St. Philips Primary School		
Ingoma Girls and Boys	Ingoma	29 August 2012
Amahubo Girls	Amahubo	29 August 2012
Soloist	Izibongo	29 August 2012
D. Impolimpoli Primary School		
Amahubo	Amahubo	31 August 2012
Afrogospel	Afrogospel	31 August 2012
Soloist	Izibongo	31 August 2012

6. TEMBE		
Entokozweni Primary School		
Ingoma Girls and Boys	Ingoma	30 August 2012
Amahubo	Amahubo	30 August 2012
Soloist	Izibongo	30 August 2012
7. ESHOWE		
Indlamu	Indlamu	2 September 2012
Eshowe HS Choir	Amahubo/Wedding Songs	2 September 2012
Starlight Band	Afrojazz	2 September 2012
8. DAKENI		
Memulo	Memulo/Amahubo	10 July 2012
9. HIMEVILLE		
A. Community Hall		
Shosholoza	Isigekle	2 June 2012
Siyanqoba	Ingoma	2 June 2012
Maskandi	Maskandi	2 June 2012
Indlamu I	Indlamu	2 June 2012
Indlamu II	Indlamu	2 June 2012
B. Himeville Anglican Church Hall		
Udumo Gospel Choir	Wedding Songs	30 June 2012
Siyanqoba	Wedding Songs	1 July 2012

Bibliography

Abrahamsen, Adele, and William Bechtel. "History and Core Themes," in *The Cambridge Handbook of Cognitive Science*, edited by Keith Frankish and William M. Ramsey, 9-28. Cambridge, Cambridge University Press, 2012.

Agawu, Kofi. *Representing African Music: Postcolonial Notes, Theories, Positions.* New York: Routledge, 2003.

______. "How We Got Out of Analysis and How to Get Back in Again," *Music Analysis* 23, nos. 2-3 (2004): 267-286.

Averill, Gage. "Introduction," in *Alan Lomax: Selected Writings 1934-1997*, edited by Ronald D. Cohen, 233-247. New York: Routledge, 2003.

Barkow, Jerome, Leda Cosmides, and John Tooby. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford, 1992.

Becker, Judith. "Ethnomusicology and Empiricism in the Twenty-First Century," *Ethnomusicology* 53, no. 3 (2009): 478-501.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann. "Language is a Complex Adaptive System: Position Paper," *Language Learning* 59, Suppl. 1 (2009): 1-26.

Blacking, John. "Indigenous Music," in *South African Music Encyclopedia*, edited by Jacques P. Malan. Pretoria: Human Sciences Research Council, 1982.

_______. *Venda Children's Songs: A Study in Ethnomusicological Analysis*. 1969, Repr. Chicago: University of Chicago Press, 1995.

Bohlman, Philip V. "Ethnomusicology," *Grove Music Online*. *Oxford Music Online*, accessed 20 April 2011.
<http://www.oxfordmusiconline.com/subscriber/article/grove/music/52178>.

Bybee, Joan. *Language usage and cognition*. Cambridge: Cambridge University Press, 2010.

Clarke, Eric, and Nicholas Cook, eds. *Empirical Musicology*. Oxford: Oxford University Press, 2004.

Cone, Edward T. *Musical form and musical performance*. New York: Norton, 1968.

Cooley, Timothy J., and Gregory Barz. "Casting Shadows: Fieldwork Is Dead! Long Live Fieldwork!" In *Shadows in the Field: New Perspectives for Fieldwork in Ethnomusicology*, Second Edition, edited by Barz and Cooley, 3-24. New York: Oxford, 2008.

Cooper, David, and Ian Sapiro. "Ethnomusicology in the Laboratory: From the Tonometer to the Digital Melograph," *Ethnomusicology Forum* 15, No. 2 (2006): 301-313.

Coplan, David. *In Township Tonight: South Africa's Black City Music and Theatre*, Second Edition. Chicago: University of Chicago Press, 2007.

Cooper, Grosvenor and Leonard B. Meyer, *The Rhythmic Structure of Music*. Chicago: University of Chicago Press, 1960.

Cowie, Fiona. "Innateness and Language," in *The Stanford Encyclopedia of Philosophy,* Summer 2010 Edition, edited by Edward N. Zalta (Stanford, CA: Stanford University), http://plato.stanford.edu/archives/sum2010/entries/innateness-language/ (accessed June 3, 2013).

Charles Darwin, *The Descent of Man*. 1871, Repr. New York: Penguin, 2004.

Chomsky, Noam. *Syntactic Structures*. The Hague: Mouton, 1957.

________. "Verbal Behavior by B.F. Skinner," *Language* 35, no. 1 (1959): 26-58.

________. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press, 1965.

________. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper and Row, 1966.

________. *Language and Mind*, First Edition. New York: Oxford, 1967.

________. *Reflections on Language*. New York, NY: Pantheon, 1975.

________. *Language and Mind*, Third Edition. Cambridge: Cambridge University Press, 2006.

________, and Morris Halle, *The Sound Pattern of English*. 1968, Reprinted Cambridge, MA: MIT Press, 1991.

Clifford, James and George E. Marcus. *Writing Culture: the poetics and politics of ethnography*. Berkeley: University of California Press, 1986.

Cook, Nicholas. “The perception of large-scale tonal closure,” *Music Perception* 5, no. 2 (1987): 197-206.

Cummins, Robert, and Denise Dellarosa Cummins, eds. *Minds, Brains, and Computers: The Foundations of Cognitive Science*. Malden, MA: Blackwell, 2000.

Cumming, Naomi. “Eugene Narmour’s Theory of Melody,” *Music Analysis* 11, nos. 2-3 (1992): 354-374.

Dissanayake, Ellen. “Antecedents of the Temporal Arts in Early Mother-Infant Interaction,” in *The Origins of Music*, edited by Nils L. Wallin, Björn Merker, and Steven Brown, 463-472. Cambridge, MA: MIT Press, 2000.

Doke, Clement M., D.M. Malcolm, J.M.A Sikakana, and B.W. Vilakazi, *English-Zulu, Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press, 2008.

Dowling, W. Jay. “Development of musical schemata in children’s spontaneous singing,” in *Cognitive processes in the perception of art*, edited by W.R. Crozier and A.J. Chapman, 145-163. Amsterdam, The Netherlands: North-Holland, 1984.

_______, and Dale L. Harwood, *Music cognition*. New York, NY: Academic Press, 1986.

_______. “The development of music perception and cognition,” in *The Psychology of Music*, Second Edition, edited by Diana Deutsch, 603-625. San Diego, CA: Academic Press, 1999.

Edworthy, Judy. “Interval and contour in melody processing,” *Music Perception* 2, No. 3 (Spring, 1985): 375-388.

Ellis, Alexander. “On the musical scales of various nations,” *Journal of the Society of the Arts* 33 (1885): 485-517.

Erlmann, Veit. *Nightsong: Performance, Power and Practice in South Africa*. Chicago: Chicago University Press, 1990.

Evans, Nicholas and Stephen C. Levinson, “The Myth of Language Universals: Language Diversity and its Importance for Cognitive Science,” *Behavioral and Brain Sciences,* 32 (2009): 429-492.

Falk, Dean. *Finding our tongues: mothers, infants and the origins of language*. New York: Basic Books, 2009.

Fitch, Tecumseh. "The biology and evolution of music: A comparative perspective," *Cognition* 100 (2006): 173-215.

Fodor, Jerry. 1983. *Modularity of Mind*. Cambridge, MA: MIT Press.

_____. *The mind doesn't work that way: The scope and limits of computational pscyhology*. Cambridge, MA: MIT Press, 2001.

Fox, Anthony J. *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals*. New York: Oxford, 2000.

Frick, Robert W. "Communicating Emotion: The Role of Prosodic Features," *Psychological Bulletin* 97, no. 3 (1985): 412-429.

Gjerdingen, Robert. *A Classic Turn of Phrase: Music and the Psychology of Convention*. Philadelphia: University of Pennsylvania Press, 1988.

_________. "Categorization of Musical Patterns by Self-Organizing Neuron-like Networks," *Music Perception* 7, No. 4 (1990): 339-370.

_________. "The Psychology of music," in *The Cambridge History of Western Music Theory*, edited by Thomas Christensen, 956-981. Cambridge: Cambridge University Press, 2008.

Godfrey-Smith, Peter. *Theory and Reality: An Introduction to the Philosophy of Science*. Chicago: University of Chicago Press, 2003.

Grauer, Victor. "Echoes of our forgotten ancestors," *Worlds of Music* 48, no. 2 (2006): 101-134.

Greenberg, Joseph H. ed. *Universals of* language. Cambridge, MA: MIT Press.

_______. "Diachrony, synchrony and language universals," in *Universals of Human Language: Vol. 1, Method and Theory*, edited by Joseph H. Greenberg, Charles A. Ferguson, & Edith A. Moravcsik, 61-92. Stanford, CA: Stanford University Press, 1978.

Goldsmith, John. *Autosegmental and Metrical Phonology*. Oxford: Blackwell, 1990.

Guy, Gregory. "Competence, performance, and the generative grammar of variation," in *Variation, change and phonological theory*, Amsterdam Studies in the theory and history of linguistic science 146, edited by Frans Hinskens, Roeland van Hout, and W. Leo Wetzels. Philadelphia: John Benjamins, 1997.

Gussenhoven, Carlos. *The Phonology of Tone and Intonation*. New York: Oxford, 2004.
Hardcastle, Charles W., J. Laver and F. Gibbon, eds. *Handbook of the Phonetic Sciences*, Second Edition. Oxford: Wiley-Blackwell, 2010.

Hart, Johan ‘t, René Collier, and Antonie Cohen. *A Perceptual Study of Intonation: An Experimental-Phonetic Approach To Speech Melody*. Cambridge: Cambridge University Press, 1990.

Hasty, Christopher. *Meter as Rhythm*. Oxford: Oxford University Press, 1997.

Hockett, Charles F. “Sound Change,” *Language* 41, no. 2 (1965): 185-204.

Hood, Mantle. *The Ethnomusicologist*. Kent, OH: Mc-Graw-Hill, 1971.

Hornbostel, Erich M. von. “African Negro Music,” *Africa: Journal of the International African Institute* 1, no. 1 (1928): 30-62.

Huron, David. “Interval-class content in equally-tempered pitch-class sets: Common scales exhibit optimum tonal consonance,” *Music Perception* 11, no. 3 (1994): 289-305.

_____. “What is Melodic Accent? Converging Evidence from Musical Practice,” *Music Perception* 13, no. 4 (1996): 489-516.

_____. “Tone and voice: A derivation of voice-leading from perceptual principles,” *Music Perception* 19, no. 1 (2001): 1-64.

_____. *Sweet Anticipation*. Cambridge, MA: MIT Press, 2006.

Huron, David, and Joy Ollen. “Agogic contrast in French and English Themes: Further Support for Patel and Daniele,” *Music Perception* 21, no. 2 (2003): 267-271.

Hutchins, Sean, J.M. Zarate, Robert J. Zatorre, and Isabelle Peretz. “An acoustical study of vocal pitch matching in congenital amusia,” *Journal of the Acoustical Society of America*, 127, no. 1 (2012): 504-512

Jackendoff, Ray. “Parallels and Nonparallels between Language and Music,” *Music Perception* 26, no. 3 (February 2009): 195-204.

Jackendoff, Ray, and Fred Lerdahl. “The Capacity for Music: What Is It, And What’s Special About It?” *Cognition* 100 (2006): 33-72.

Jakobson, Roman C., Gunnar Fant and Morris Halle, *Preliminaries to Speech Analysis. The distinctive features and their correlates*. Cambridge, MA: MIT Press, 1952.

Jones, A.M. *Studies in African Music*. New York: Oxford University Press, 1959.

Joseph, Rosemary. "Zulu women's Bow Songs: Ruminations on Love," *Bulletin of the School of Oriental and African Studies*, *University of London* 50, No. 1, (1987): 90-119.

Justus, Timothy C., Jamshed J. Bharucha, "Music Perception and Cognition," in *Stevens' Handbook of Experimental Psychology*, Third Edition, edited by S. Yantis and H. Pashler. New York: Wiley, 2002.

Kirby, Percival. *The musical instruments of the native races of South Africa*, Second Edition. Johannesburg: Witwatersrand University Press, 1965.

Krumhansl, Carol. *Cognitive Foundations of Musical Pitch*, Second Edition. New York: Oxford, 2001.

_________, and Edward J. Kessler. "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys," *Psychological Review* 89 (1982): 334-368.

Kubik, Gerhard. *Theory of African Music*, Volume 2. Chicago: Chicago University Press, 2010.

Ladd, D. Robert. *Intonational Phonology*, Second Edition. Cambridge: Cambridge University Press, 2008.

____. "An Integrated View of Phonetics, Phonology, and Prosody," in *Language, Music, and the Brain: A Mysterious Relationship*, edited by Michael Arbib, 273-288. Cambridge, MA: MIT Press, 2013.

Laver, John. *Principles of phonetics*. Cambridge University Press, 1994.

Lerdahl, Fred. *Tonal Pitch Space*. New York: Oxford, 2001.

______. "Genesis and Architecture of the GTTM Project," *Music Perception* 26, No. 3 (2009): 187-194.

______, and Ray Jackendoff, *A Generative Theory of Tonal Music*. New York: Oxford, 1983.

Levitin, Daniel, and Anna Tirovolas. "Current advances in the cognitive neuroscience of music," *Annals of the New York Academy of Sciences* 1156, Iss. 1 (March 2009): 211-231.

Liberman, Mark, and Alan Prince. "On Stress and Linguistic Rhythm," *Linguistic Inquiry* 8, no. 2 (1977): 249-336.

Liu, Fang, Aniruddh D. Patel, Adrian Fourcin, and Lauren Stewart, "Intonation processing in congenital amusia: discrimination, identification and imitation," *Brain* 133, No. 6 (April 2010): 1682-1693.

London, Justin. *Hearing in Time: Psychological Aspects of Musical Meter*. New York: Oxford, 2004.

Maggs, Tim. "The Iron Age Farming Communities," in *Natal and Zululand From Earliest Times to 1910: A New History*, edited by Andrew Duminy and Bill Guest, 28-48. Pietermaritzburg: University of Natal Press/Shuter and Shooter.

Manning, Christopher D., and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press, 1999.

Marvin, Elizabeth West, and Paul A. Laprade. "Relating Musical Contours: Extensions of a Theory for Contour," *Journal of Music Theory* 31, No. 2 (1987): 225-267.

McClelland, James L., and Joan Bybee, "Gradience of Gradience: A reply to Jackendoff," *The Linguistic Review* 24: (2007), 437-355.

McCreless, Patrick. "Syntagmatics and Paradigmatics: Some Implications for the Analysis of Chromaticism in Tonal Music," *Music Theory Spectrum* 13, no. 2 (1991): 147-178.

Meintjes, Louise. "Shoot the Sargeant, Shatter the Mountain: The Production of Masculinity in Zulu Ngoma Song and Dance in Post-Apartheid South Africa," *Ethnomusicology Forum* 13, no. 2 (2004): 173-201.

Merriam, Alan. *The Anthropology of Music*. Evanston, IL: Northwestern University Press, 1964.

Meyer, Leonard. *Emotion and Meaning in Music*. Chicago: University of Chicago Press, 1956.

_____. "Universalism and Relativism in the Study of Ethnic Music," *Ethnomusicology* 4, No. 2, (1960): 49-54.

_____. *Explaining Music: Essays and Explorations*. Chicago: Chicago University Press, 1973.

_____. “A universe of universals,” *Journal of Musicology* 16, no. 1 (1998): 3-25.

Mthethwa, Bongani. *Zulu Folksongs: History, Nature and Classroom Potential*. BMUS dissertation: University of Natal, 1979.
Muller, Carol A. *Focus: Music of South Africa*, Second Edition. New York: Routledge, 2004.

Nan, Yun, Yanan Sun, and Isabelle Peretz. “Congenital amusia in speakers of a tone language: association with lexical tone agnosia,” *Brain* 133, no. 9 (August 2010), 2635-2642.

Narmour, Eugene. *Beyond Schenkerism: The Need for Alternatives in Music Analysis*. Chicago: University of Chicago Press, 1977.

_______. “Some Major Theoretical Problems Concerning the Concept of Hierarchy in the Analysis of Tonal Music,” *Music Perception* 1, no. 2 (1983): 129.

_______. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press, 1990.

_______. *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago: University of Chicago Press, 1992.

Nettl, Bruno. “Comparative Study and Comparative Musicology: Comments on Disciplinary History,” in *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology*, edited by Rolf Bader, Christiane Neuhaus, and Ulrich Morgernstern. Frankfurt: Peter Lang, 2010.

Neuhaus, Christiane. “Processing musical form: Behavioral and neurocognitive approaches. *Musicae Scientiae* 17, no. 1 (2013): 109-127.

Nketia, J.H. Kwabena. *The Music of Africa*. New York: Norton, 1974.

Nooteboom, Sieb. “The Prosody of Speech: Melody and Rhythm,” in *Handbook of the Phonetic Sciences* edited by W. Hardcastle and J. Laver, 640-673. Oxford: Blackwell, 1997.

Palmer, Caroline, and Sean Hutchins, “What is musical prosody?” In *The Psychology of Learning and Motivation, Vol. 46: Advances in Research and Theory*, edited by Barry Ross, 245-278. San Diego, CA: Academic Press, 2006.

Patel, Aniruddh D. *Music, language and the brain*. New York: Oxford, 2008.

____, John R. Iversen, and Jason C. Rosenberg. "Comparing the rhythm and melody of speech and music: The case of British English and French," *Journal of the Acoustical Society of America* 119 (2006): 3034-3047.

____. "A new approach to the cognitive neuroscience of melody," in *The Cognitive Neuroscience of Music*, edited by Isabelle Peretz and Robert Zatorre, 326-345. New York: Oxford, 2003.

____, Meredith Wong, Jessica Foxton, Aliette Lochy, and Isabelle Peretz. "Speech intonation perception deficits in musical tone deafness (congenital amusia)," *Music Perception* 25, no. 4 (2008): 357-368.

____, and Stephen Demorest, "Comparative Music Cognition: Cross-Species and Cross-Cultural Studies," in *The Psychology of Music*, Third Edition, edited by Diana Deutsch, 647-682. London: Academic Press, 2012.

Peretz, Isabelle, Anne Sophie Champod, and Krista Hyde. "Varieties of Musical Disorders: The Montreal Battery of Evaluation of Amusia," *Annals of the New York Academy of Sciences*, 999 (November 2003): 58-75.

_____. "Music, Language and Modularity Framed in Action," *Psychologica Belgica* 49, nos. 2-3 (2009): 159-160.

_____. "The biological foundations of music: Insights from congenital amusia," in *The Psychology of Music*, Third Edition, edited by Diana Deutsch, 551-564. San Diego, CA: Academic Press, 2012.

_____, and Robert J. Zatorre, "Brain organization for music processing," *Annual Review of Psychology*, 56 (February 2005): 89-114.

Pierrehumbert, Janet. *The Phonology and Phonetics of English Intonation*. PhD: MIT, 1980.

___________. "What people know about the sounds of language," *Linguistic Sciences* 29 (2000): 111-120.

Pike, Kenneth L. *Tone Languages: A technique for determining the number and type of pitch contrasts in a language, with studies in tonemic substitution and fusion*. Ann Arbor: University of Michigan Press, 1948.

Pinker, Steven, and Paul Bloom. "Natural Language and Natural Selection," *Behavioral and Brain Sciences* 13 (1990): 707-784.

Pooley, Thomas M. "Dismantling music: Reductionist models and evolutionary explanations in music cognition," in *Topicality of Musical Universals/Actualité des Universaux Musicaux*, edited by Jean-Luc Leroy, 43-48. Paris: Éditions des archives contemporaines, 2013.

Port, Robert F., and Adam P. Leary, "Against Formal Phonology," *Language* 81, no. 4 (2005): 927-964.

Quinn, Ian. "The combinatorial model of pitch contour," *Music Perception* 16, no. 4 (1999): 439-456.

Randel, Don M., ed. "Melody [fr. Lat. Melodia, Fr. Gr. Meloidia, Fr. Melos]." *The Harvard Dictionary of Music*. Cambridge: Harvard UP, 2003. *Credo Reference*. 25 June 2004. Web. 5 Nov. 2013. <http://0www.credoreference.com.oasis.unisa.ac.za/entry/harvdictmusic/melody_fr_lat_melodia_fr_gr_meloidia_fr_melos>.

Rosner, Burton, and Leonard B. Meyer, "The Perceptual Roles of Melodic Process, Contour, and Form." *Music Perception* 3, no. 1, (1986): 1-39.

Ruwet, Nicolas, and Mark Everist, "Methods of Analysis in Musicology," *Music Analysis* 6, no. 1/2, (1987): 3-9+11-36.

Rycroft, David. "A Royal Account of Music in Zulu Life with Translation, Annotation, and Musical Transcription," *Bulletin of the School of Oriental and African Studies, University of London* 38, No. 2 (1975): 351-402.

______. "Indigenous Music," in *South African Music Encylopedia*, ed. Jacques P. Malan. Pretoria: Human Sciences Research Council, 1982.

Saussure, Ferdinand de. *Course in General Linguistics*, translated by Wade Baskin, edited by Perry Meisel and Haun Saussy. New York: Columbia University Press, 2011.

Schmuckler, Mark A. "Pitch and Pitch Structures," in *Ecological Psychoacoustics*, edited by J. Neuhoff, 271-315. San Diego, CA: Academic Press, 2004.

___________. "Components of melodic processing," in *Oxford Handbook of Music Psychology*, edited by Susan Hallam, Ian Cross, and Michael Thaut, 93-106. Oxford: Oxford University Press, 2009.

__________. “Melodic Contour Similarity Using Folk Melodies,” *Music Perception: An Interdisciplinary Journal* 28, No. 2, (2010): 169-194.

__________, and Robert Tomovski. “Perceptual Tests of an Algorithm for Musical Key-Finding,” *Journal of Experimental Psychology: Human Perception and Performance* 31, No. 5 (2005): 1124-1149.

Schneider, Albrecht. “Comparative and Systematic Musicology in Relation to Ethnomusicology: A Historical and Methodological Survey,” Ethnomusicology 50, No. 2 (2006): 236-258.

Schön, Daniele, Reyna Leigh Gordon, and Mireille Besson. “Musical and Linguistic Processing in Song Perception,” The Neurosciences and Music II: From Perception to Performance, *Annals of the New York Academy of Sciences* 1060, (December 2005): 71-81.

Schubert, Emery. “Modeling Perceived Emotion With Continuous Musical Features,” *Music Perception* 21, No. 4 (2004): 561-585.

Steingo, Gavin. *After Apartheid: Kwaito Music and the Aesthetics of Freedom*. PhD Dissertation: University of Pennsylvania, 2010.

Stevens, Catherine. “Music Perception and Cognition: A review of recent cross-cultural research,” *Topics in Cognitive Science*, 4 (2012): 653-654.

Tagg, Philip. “Analysing Popular Music: Theory, Method and Practice,” *Popular Music* 2 (1982): 37-67.

Tillmann, Barbara B. and Emmanuel Bigand, “The Relative Importance of Local and Global Structures in Music Perception,” Special Issue: Art, Mind, and Cognitive Science, *The Journal of Aesthetics and Art Criticism* 62, No. 2, (2004): 211-222.

_____, Dennis Burnham, Sebastien Nguyen, Nicolas Grimault, Natalie Gosselin, and Isabelle Peretz. “Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages,” *Frontiers in Psychology*, 2 (June 2011): 1-15.

Titon, Jeff Todd. “Ecology, Phenomenology, and Biocultural Thinking: A Response to Judith Becker,” *Ethnomusicology* 53, no. 3 (2009): 502-509.

Tomasello, Michael. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press, 2003.

________, *Origins of Human Communication*. Cambridge, MA: MIT Press, 2010.

Tomlinson, Gary. “The Web of Culture: A Context for Musicology,” *19th-Century Music* 7, No. 3, Essays for Joseph Kerman (1984): 350-362.

_________. “Evolutionary Studies in the Humanities: The Case of Music,” *Critical Inquiry* 39 (Summer 2013): 647-675.

Trehub, Sandra. “Musical predispositions in infancy,” *Annals of the New York Academy of Sciences*, 930, Iss. 1 (2001): 1-16.

______, “Communication, Music, and Language in Infancy,”. In *Language, Music and the Brain: A Mysterious Relationship*, edited by Michael Arbib, 463-479. Cambridge, MA: MIT Press, 2013.

Allan Vurma and Jaan Ross, “Priorities in Voice Training: Carrying Power or Tone Quality,” *Musicae Scientiae* 4 (Spring 2000): 75-93.

Waugh, Linda R., and Cornelis H. van Schooneveld, eds. *The Melody of Language: Prosodics and Intonation*. Baltimore, MD: University Park Press, 1980.

Webster, James. “*Formenlehre* in Theory and Practice,” in *Musical Form, Forms, and Formenlehre: Three Methodological Reflections*, edited by Pieter Bergé, 123-139. Leuven: Leuven University Press, 2009.

Weiskopf, Daniel A. “A critical review of Jerry A. Fodor’s *The mind doesn’t work that way*,” *Philosophical Psychology* 15, No. 4 (2002): 555.

Wennerstrom, Ann. *The Music of Everyday Speech: Prosody and Discourse Analysis.* Oxford: Oxford University Press, 2001.

Wheatley, Susan. *An application of chunking to the memory and performance of melodic patterns*. PhD dissertation: University of Michigan, 1991.

Whitelaw, Gavin. “A Brief Archaeology of Precolonial Farming in KwaZulu-Natal,” in *Zulu Identities: Being Zulu, Past and Present*, edited by Benedict Carton, John Laband and Jabulani Sithole, 47-61. Pietermaritzburg: University of KwaZulu-Natal Press, 2008.

Wright, John. “Reflections on the Politics of Being ‘Zulu,’” in *Zulu Identities: Being Zulu, Past and Present*, edited by Benedict Carton, John Laband and Jabulani Sithole, 35-44. Pietermartizburg: University of KwaZulu-Natal Press, 2008.

______, and Carolyn Hamilton. “Traditions and transformations: The Phongolo-Mzimkhulu region in the late eighteenth and early nineteenth centuries,” in *Zulu*

Identities: Being Zulu, Past and Present, edited by Benedict Carton, John Laband and Jabulani Sithole, 69-82. Pietermaritzburg: University of KwaZulu-Natal Press, 2008.

Yip, Moira. *Tone*. Cambridge: Cambridge University Press, 2002.

Zbikowski, Lawrence. *Conceptualizing Music: Cognitive structure, theory and analysis*. AMS Studies in Music. New York: Oxford, 2003.

Discography

Rimes, LeAnn. *You Light Up My Life: Inspirational Songs*. © 1997 by Curb Records. 77885. Compact disc.

Karajan, Herbert von. *Beethoven: Nine Symphonies*. With Berlin Philharmonic Orchestra, Gundula Janowitz, Waldemar Kmentt, Hilde Rossel-Majdan, Walter Berry, Wiener Singverein. © 2007, 1963 by Deutsche Grammaphon. DG 429 036. Compact disc.