THE UNIVERSITY OF CHICAGO

ORGANIZING THE SOUND OF THE VOICE:

WESTERN MUSIC'S RELATIONSHIP WITH RECORDED SPEECH, 1965–2020

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE HUMANITIES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF MUSIC

BY

REBECCA FLORE

CHICAGO, ILLINOIS

JUNE 2021

TABLE OF CONTENTS

ABSTRACT

Throughout Western music from the 1960s until today—in genres and epochs ranging from concert music, to experimental albums in popular music, to goofy YouTube videos—we can hear a peculiar phenomenon: musical works in which the music is organized around sound recordings of speech. Works like these present an interesting opportunity to return to the age old question, what does music have in common with natural language? Furthermore, how can the two competing organization systems of sound found in natural language and music work together? How do composers and musicians deal with conflicts between the resulting sounds of each organizational system? What is the role of the speaker and the composer and musicians in the work of authorship in such music? This dissertation seeks to answer these questions and in doing so it will dig deeper into the relationship between sound, music, and its human agents in modern Western music.

I'm particularly interested in music that relies on recordings of speech because these recordings allow for a repeatability, a set of fixed moments in pitch and time that, I propose, these musicians and composers analyze musically in order to sort them into the musical structure. This dissertation has two major parts. In the first, I survey the music, across genres and art vs. popular distinctions, and uncover particular compositional strategies that composers and musicians use to musicalize speech. In the second part, I take a closer look at individual approaches to the musicalization of speech, exploring 1) the role of the speaker as a collaborator—distanced by time and space—in creating the structure of the musical work, and 2) the impact of outside influences on the composers, from speakers to other composers to musical styles, on the resulting musical works.

ACKNOWLEDGEMENTS

First and foremost, thank you to my advisor Larry Zbikowski. Larry diligently read drafts, helped me refine my ideas, pushed me in new directions, fixed my typos, encouraged me to develop my writing style, supervised my teaching, gave me career advice, etc. No matter how uncertain I was about this project or my professional life, I always left meetings with him feeling better and with a new sense of direction. I'm eternally grateful for his mentorship and support during my time at UChicago.

Thank you to my committee members Steve Rings and Jennifer Iverson, who provided much-needed different perspectives on my research, were always enthusiastic about my project, and encouraged me. Thanks also to Seth Brodsky, Robert L. Kendrick, and Robert Hasegawa, who served as sounding boards for my early thoughts about this project.

I want to give my deepest thanks my colleagues who have read drafts, listened to me present, given feedback, helped me work through ideas, and/or raised questions: George Adams, Devon Borowski, Jon Bullock, Barbara Dietlinger, Anna Gatdula, Joshua Klopfenstein, John Lawrence, Ailsa Lipscombe, Nancy Murphy, Hannah Rogers, Siavash Sabetrohani, Max Silva, Florian Walch, and Andrew White—and apologies to anyone I've accidentally omitted! This dissertation would be much poorer without all of your feedback and support.

Thank you to the Hanna Holborn Gray Special Collections Research Center at University of Chicago's Regenstein Library for being a great place to work for almost six years during my time here. I'm forever indebted to Elizabeth Frengel and Catherine Uecker for their mentorship and their encouragement to explore academic library work as a career.

To my parents: I couldn't have gotten so far without your love and support. To Joey Hribar: thanks for being there for the whole grad school odyssey and listening to me sigh whenever a thought about this dissertation popped into my head outside of work hours.

I appreciate you all.

TABLE OF FIGURES

Chapter 1: Introduction

Imagine you're surfing YouTube and come across the video "Ants in My Eyes Johnson w/Drums."[1] The video features drummer David Dockery, a personality with many videos on YouTube, playing along to a spoken monologue from the cartoon *Rick and Morty* in which the titular electronics salesman tries to hawk his wares in a commercial. You're drawn to the interaction between the drums and the spoken words. From the drums you hear a complex rhythmic/metric surface, which features clear rhythmic patterns over near-constant shifts in meter. These shifts seem to frame the sound of the speech in some way. From the voice, you hear a frenetic delivery of the monologue—it sounds like the spoken component was meant to be dramatic but never intended to be musical. Nonetheless, the words seem to take on musical rhythmic qualities when set with the drums.

In all likelihood, this isn't the first time you've encountered a musical work that uses recorded speech as some sort of structuring agent. These works are not limited to one style, school, or genre; the recorded speaking voice as compositional object proliferates throughout the late twentieth and early twenty-first century Western musical world, and can be found in minimalist, spectral, electronic, jazz, and hip hop works. The pieces we will explore in this dissertation cover a wide range of musical styles in order to investigate the high-level, trans-genre phenomena of structuring music around recorded speech.

The "Ants in My Eyes Johnson w/Drums" example brings up many questions that are central to this dissertation and will apply across this diverse corpus of music (from experimental popular music to music of modernist or postmodernist traditions to internet memes): How similar

[1] David Dockery, "Ants in My Eyes Johnson w/Drums," YouTube, July 26, 2017, https://www.youtube.com/watch?v=G8UT6nkSWT4.

is music to speech? Can speech tip over into music only in selected instances, or is this kind of slippage simply a feature of human cognition? If speech and music are often associated with discrete units of auditory processing, how does music adapt when confronted with spoken language as a compositional object? How do music creators—composers, songwriters, producers, and musicians—use spoken language in the production of cultural products (musical works) that do not employ language as their primary communicative resource, and how do listeners apprehend this use?

*****

In order to consider the musicalization of speech, we must first consider our source "text"—the original recording of the scene, a clip from *Rick and Morty*'s eighth episode of season 1, "Rixty Minutes."[2] A deeper dive into this clip will help us get a better sense of how the speech is delivered and what it means. In this episode, scientist Rick hacks a cable box to allow himself and the Smith family to watch TV programming from alternate universes. As the rest of the family leaves the room, Rick and Morty (the youngest child of the Smith family) remain on the couch and focus their attention on the TV. The perspective shifts to view the TV from behind Rick and Morty's vantage point on the couch so that the backs of their heads are still in the scene, underscoring the diegetic nature of the TV programming within the larger plot (Figure 1.1). "Ants in My Eyes Johnson Electronics" is the first of many surreal, alternate-universe advertisements in the episode, all of which were improvised by *Rick and Morty* co-creator Justin Roiland. This improvisation might account for the jerky, frequently changing expressiveness of Ants in My Eyes Johnson that makes this clip so ripe for Dockery's musicalization.

[2] This clip can be found on Adult Swim's YouTube channel excerpted from the larger episode. Adult Swim, "Ants In My Eyes Johnson | Rick and Morty | Adult Swim," YouTube, July 22, 2015, https://www.youtube.com/watch?v=G4BkGJj5BXg.

Figure 1.1. Adult Swim, "Ants In My Eyes Johnson | Rick and Morty | Adult Swim," YouTube, July 22, 2015, https://www.youtube.com/watch?v=G4BkGJj5BXg, 0:06

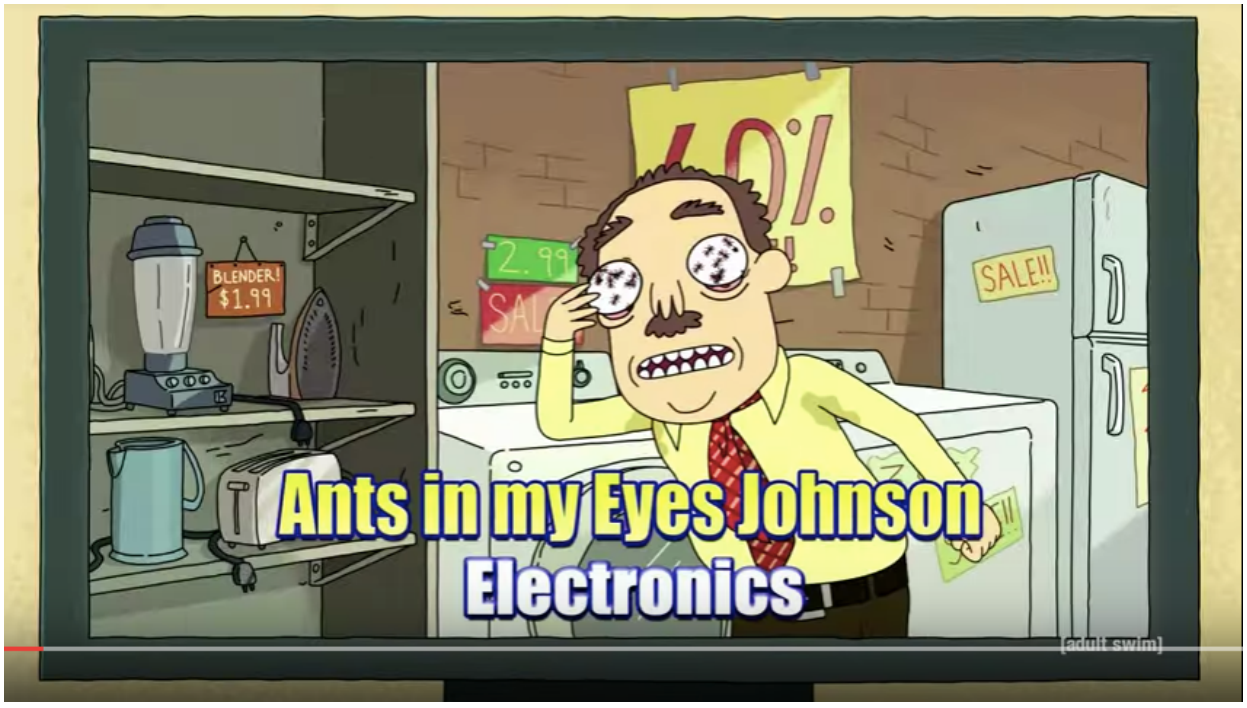
Dockery is not the only person who must contend with organizing the ad-libbed quality of the voice into some sort of structure: sticking to the *Ricky and Morty* clip itself for a moment, the animators had to link the narrative that Roiland improvised to visuals. According to Daniel Kurland, a reviewer for Vulture, "the animators were given free reign [sic] to do whatever they wanted with the nonsense that's being spouted and it amounts to a prime example of how funny the *animators* (or art department) of a series can be, in addition to the writers. This is as much a showcase for them as it is the writing staff."[3]

The off-the-cuff nature of the monologue and the animation also extends to the larger soundscape of the alternate-universe ads. There is more to the soundworld of the commercial

[3] Daniel Kurland, "'Rick and Morty' Threw Out the Script to Make One of Their Best Episodes," Vulture, May 24, 2016, http://www.vulture.com/2016/05/rick-and-morty-threw-out-the-script-to-make-one-of-their-best-episodes.html.

than merely the speaking voice, sounds that are perhaps less notable in Dockery's musicalization as they are barely perceptible underneath the drums. When the frame narrows to cut out Rick and Morty from the shot (0:11, shown in Figure 1.2), the voice becomes clearer, drawing our focus to the alternate universe of the ad and away from Rick and Morty's universe. Likewise, when we zoom out to return to our original vantage point behind the couch (4:22), the voice returns to its original tinny, faraway quality. One can hear ants crawling throughout, and quiet background music plays within the commercial. As products and prices pop up onscreen, they're accompanied by dinging noises (0:12–14, 0:23–26). There is also one emphatic "boing!" when cutting to a close-up of Ants in My Eyes's face at 0:18. We hear crackling noises as he catches on fire at 0:33. And as customers swoop in to take away products, there are shuffling noises (0:34–44). Some of these noises—the exclamatory dings and boings in particular—are emphasized in Dockery's drumming by coinciding with heavily accented beats.

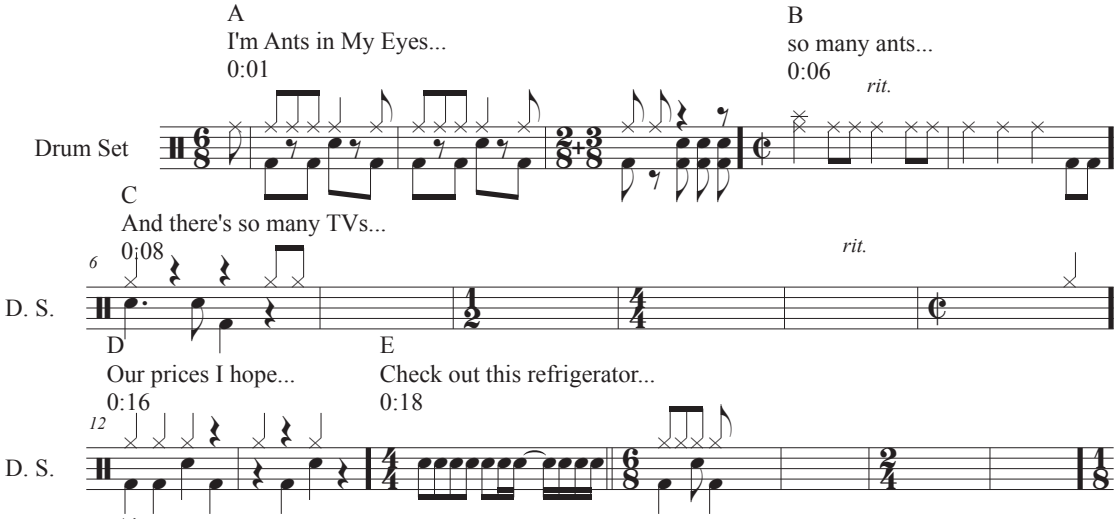
Figure 1.2. Adult Swim, "Ants In My Eyes Johnson | Rick and Morty | Adult Swim," YouTube, July 22, 2015, https://www.youtube.com/watch?v=G4BkGJj5BXg, 0:11



*****

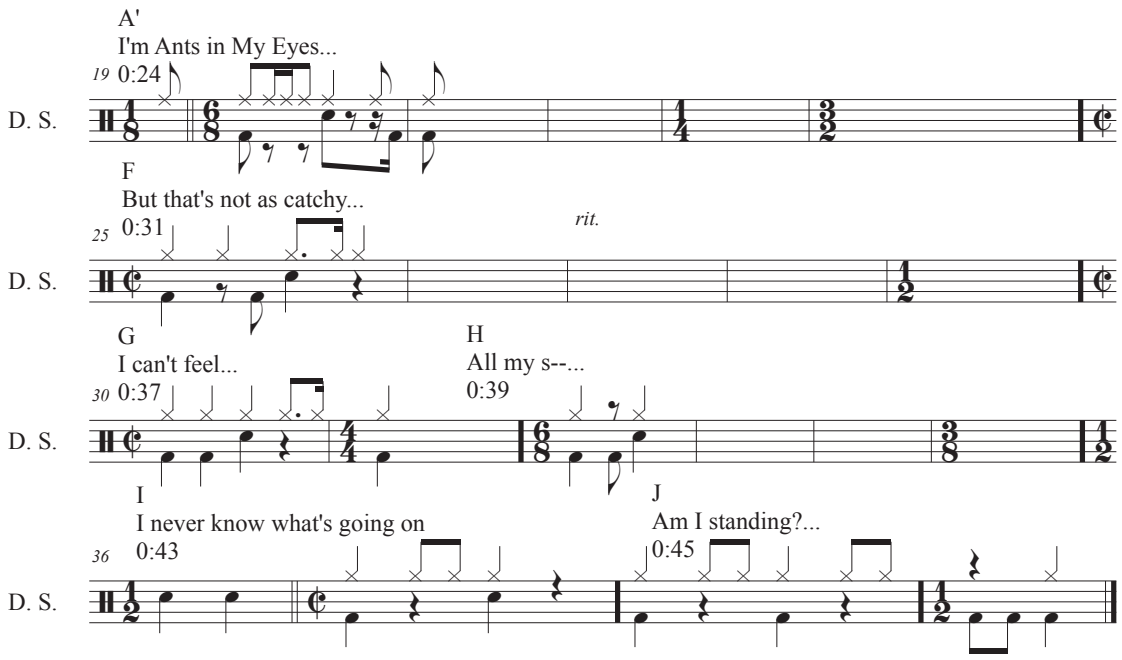
Should you watch "Ants in My Eyes Johnson w/Drums" again after acquainting yourself with the source material, you may begin to notice musical structures beyond merely the rhythmic patterns. The salesman introduces himself twice (0:02–0:05 and 0:25–0:28), and each time the spoken phrase «I'm Ants in My Eyes Johnson» is accompanied by the same compound-meter drum pattern.[4] (See Figure 1.3 for a diagram of the form with meter changes.) This repeated motif reinforces some basic sense of form. The repetition of the spoken phrase and rhythmic motif divide the musical work into two large sections, creating a sense of "starting over" about midway through the clip when this occurs the second time. The semantic content of the speech switches its focus at this middle point; what falls into the "first verse" is mostly about selling electronics, and the "second verse" is mostly about the salesman's health problems. All this to say, the musical structure can be used to create or underscore formal structures that aren't immediately apparent in speech alone.

Figure 1.3. Form diagram, Dockery, "Ants in My Eyes Johnson w/Drums"



[4] I will be using European quotation marks (« ») throughout this dissertation to denote recorded speech as a way of differentiating such speech from text that is typically expressed with standard quotation marks, such as denoting citations or titles.

Figure 1.3, continued. Form diagram, Dockery, "Ants in My Eyes Johnson w/Drums"



This musical formal division informed by semantic content occurs not only on a large-scale sectional level of the musical form but also on a phrase level. Phrase divisions that are signalled by changes in drum pattern are marked in Figure 1.3 by thick black measure lines. As we can see, these divisions vary wildly in hypermetrical duration but are united in that they correspond with the beginning of spoken phrases. For reference, Figure 1.4 shows a transcription of the speech divided into sections that correspond with musical phrases. These phrases sometimes correspond with only one spoken sentence but at other times correspond with several sentences that elaborate a single theme. Some are underscored by just one drum groove while others contain several. On some level, the drum patterns may mimic the expressive quality of the voice. For instance, under «All my s—all my nerves, they don't allow for the sensation of touch» (0:39–0:43), we hear the drums play a compound grouping of 6/8. This grouping is set up by the division of the first two spoken phrases, «All my s—» and «all my nerves», into three-syllable units in which each syllable is of roughly equal length. In the preceding phrase, the short-short-

long syllable pattern of «I can't feel» helps lay the groundwork for a 4/4 feel. We can see that Dockery's choice of meters, drum patterns, and phrasing are carefully molded to the temporal scaffolding prepared by the speech on both a semantic and phonemic level, and that these decisions provide the basis for musical form in this clip.

Figure 1.4. Transcription of speech grouped by musical formal section, Dockery, "Ants in My Eyes Johnson w/Drums"

A. I'm Ants in My Eyes Johnson, here at Ants in My Eyes Johnson's Electronics.

B. I mean, there's so many ants in my eyes.

C. And there's so many TVs, microwaves, radios—I think. I can't—I'm not a hundred percent sure what we have here in stock because I can't see anything.

D. Our prices I hope aren't too low.

E. Check out this refrigerator. Only two hundred dollars. What about this microwave? Only a hundred dollars. That's fair.

A'. I'm Ants in My Eyes Johnson. Everything's black. I can't see a thing, and also I can't feel anything either. Did I mention that?

F. But that's not as catchy as having ants in your eyes so that always goes, you know, off by the wayside.

G. I can't feel. It's a very rare disease.

H. All my s—all my nerves, they don't allow for the sensation of touch.

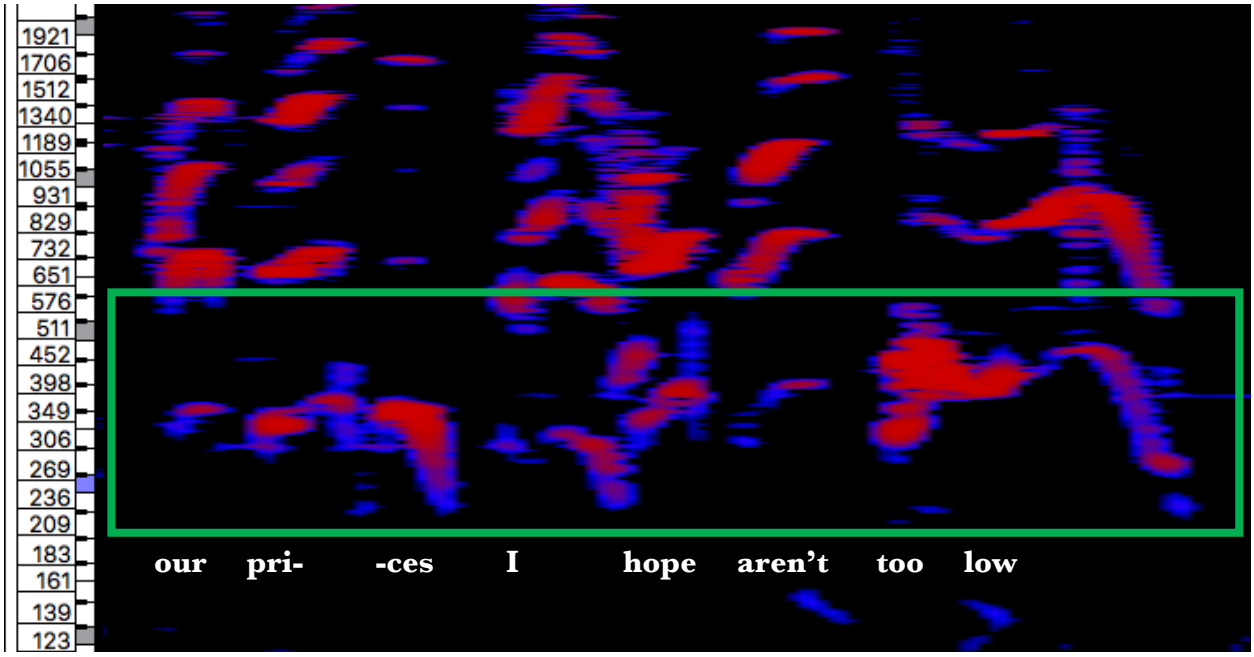
I. I never know what's going on.

J. Am I standing? Sitting? I don't know.

As you listen, you may also begin to pay attention to the pitch of the speaking voice. The voice is highly expressive, frantic even, often drifting up in pitch over the course of a sentence so that the final exclamation feels climactic. Take for example the phrase «Our prices I hope aren't too low». Figure 1.5 shows a melodic range spectrogram of this phrase from the cartoon clip itself

sans drums.[5] Onsets of syllables are indicated by their written equivalent and the first formant is outlined in green.[6] As we can see, the onsets of the syllables remain roughly steady in «our prices» and then gradually climb until reaching an apex midway through the final syllable «low». The keyboard visualization on the left side of the spectrogram helps us see that this gesture covers a fairly wide pitch range. The inflection of this phrase becomes important when transliterated to a musical imitation of the passage.

Figure 1.5. Spectrogram of «Our prices I hope aren't too low», Adult Swim, "Ants In My Eyes Johnson | Rick and Morty | Adult Swim," 0:19–0:22

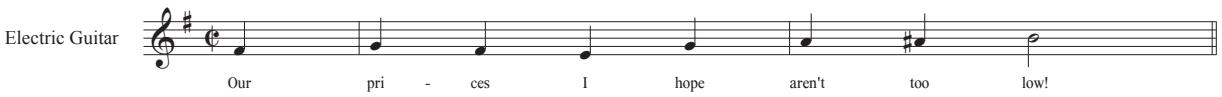


[5] Adult Swim, "Ants In My Eyes Johnson | Rick and Morty | Adult Swim," YouTube, July 22, 2015, https://www.youtube.com/watch?v=G4BkGJj5BXg.

[6] Formants are resonances of the vocal tract produced when a speaker utters vowel sounds. Formant frequencies are the center of the range of frequencies in a formant. The frequencies of the first three formants are measured to explore differences between vowel sounds. See Jody Kreiman and Diana Sidtis, *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception* (Malden, MA: Wiley-Blackwell, 2011), 52–54.

Were you to survey some of the videos on YouTube that build on Dockery's work you would find that others have noticed how musical the pitch of the speaking voice sounds in this context, for there are several videos in which instrumentalists further musicalize the clip. For instance, guitarist Marc Anthony Figueras added two further tracks to Dockery's: a melody part and a rhythm guitar part, which together provide a harmonic structure for the piece.[7] In the description for the video, Figueras notes that the "melody follows the pitch and inflection [of the voice], not perfectly but close." We can see this demonstrated in Figueras's rendering of «Our prices I hope aren't too low», a transcription of which is shown in Figure 1.6.

Figure 1.6. Transcription of melody guitar realization of «Our prices I hope aren't too low», Figueras, "'Ants In My Eyes Johnson' with Guitar," 0:16–0:18



Figueras's musical interpretation of the pitch and rhythmic qualities of speech (supported by Dockery's drum patterns) is of course not the only option for musicalizing speech. When we consider only the acoustical features of speech, setting aside their semantic meanings, they take on an unpredictable, aleatoric quality, one that the musician must confront in order to fit them into a given compositional system. Although both speech and music have syntactic, pitch, and temporal components, these do not map directly onto one another and are indeed often at odds with one another. Recorded vocal utterances have their own pitch, rhythmic, formal, and syntactic content, and composers must develop strategies to adapt or integrate speech into musical frameworks.

[7] Marc Anthony Figueras, "'Ants In My Eyes Johnson' with Guitar," YouTube, August 11, 2017, https://www.youtube.com/watch?v=QBgjsm0XQ4Y.

Because musical frameworks will be important throughout this dissertation, let me take a moment to explain them. As the notion was developed by Lawrence Zbikowski, such frameworks are resources that musicians use to shape their utterances.[8] Temporal frameworks provide a means to evaluate the "when" of musical events. Meter, for instance, provides a relatively immediate temporal framework. Once a temporal framework has been established, listeners can understand whether a musical event is early, or late, or right on time. Tonal frameworks provide a means to evaluate the "where" in pitch space of musical events, and can be used to shape listening over longer spans of time. Once a tonal framework has been established, listeners can hear a return to the same point in pitch space—the tonic or the dominant, for instance—as well as directional relationships within pitch space. Both of these resources will be important as we reckon with how composers have taken naturally-occurring sounds that may not conform on their own to temporal or tonal frameworks—like the sounds of speech—and incorporated them into musical works.

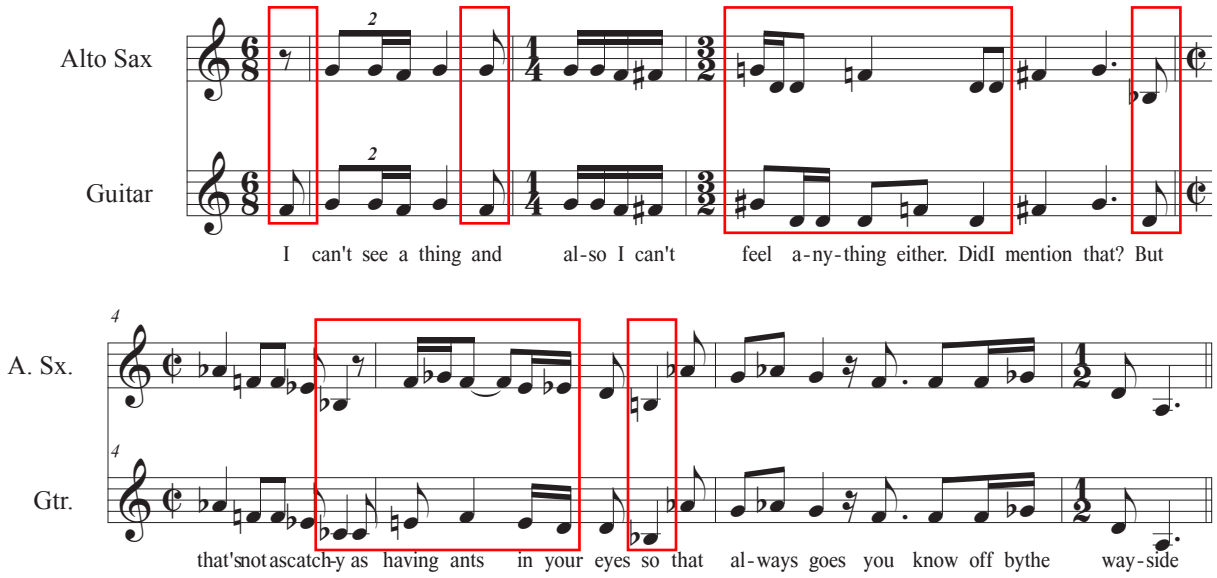
Let's now return to the world of Rick and Morty. You decide to watch another video, one that has built on Figueras's modification of Dockery's original. In "The Rick and Morty Band | Ants in My Eyes Johnson Cover on Drums, Guitar, Bass, and Sax," two more tracks have been added to Figueras's version: an electric bass and an alto saxophone.[9] Saxophonist Sancheneering's melody largely follows Figueras's melody guitar part—and why not? Figueras set out a musicalization of the speech recording that makes sense within the pitch framework set

[8] For more on musical frameworks, see Lawrence M. Zbikowski, *Foundations of Musical Grammar*, Oxford Studies in Music Theory (New York: Oxford University Press, 2017), 128.
[9] Sancheneering, "The Rick and Morty Band | Ants in My Eyes Johnson Cover on Drums, Guitar, Bass, and Sax," YouTube, September 7, 2017, https://www.youtube.com/watch?v=Pvidk3w4MfA.

up by his rhythm guitar track—but at times Sancheneering's saxophone melody pulls away from Figueras's guitar melody, bringing to light the different possibilities for musicalizing speech.

Figure 1.7. Transcription of guitar and alto saxophone melody corresponding to « I can't see a thing, and also I can't feel anything either. Did I mention that? But that's not as catchy as having ants in your eyes so that always goes, you know, off by the wayside », Sancheneering, "The Rick and Morty Band | Ants in My Eyes Johnson Cover on Drums, Guitar, Bass, and Sax," 0:27–0:37

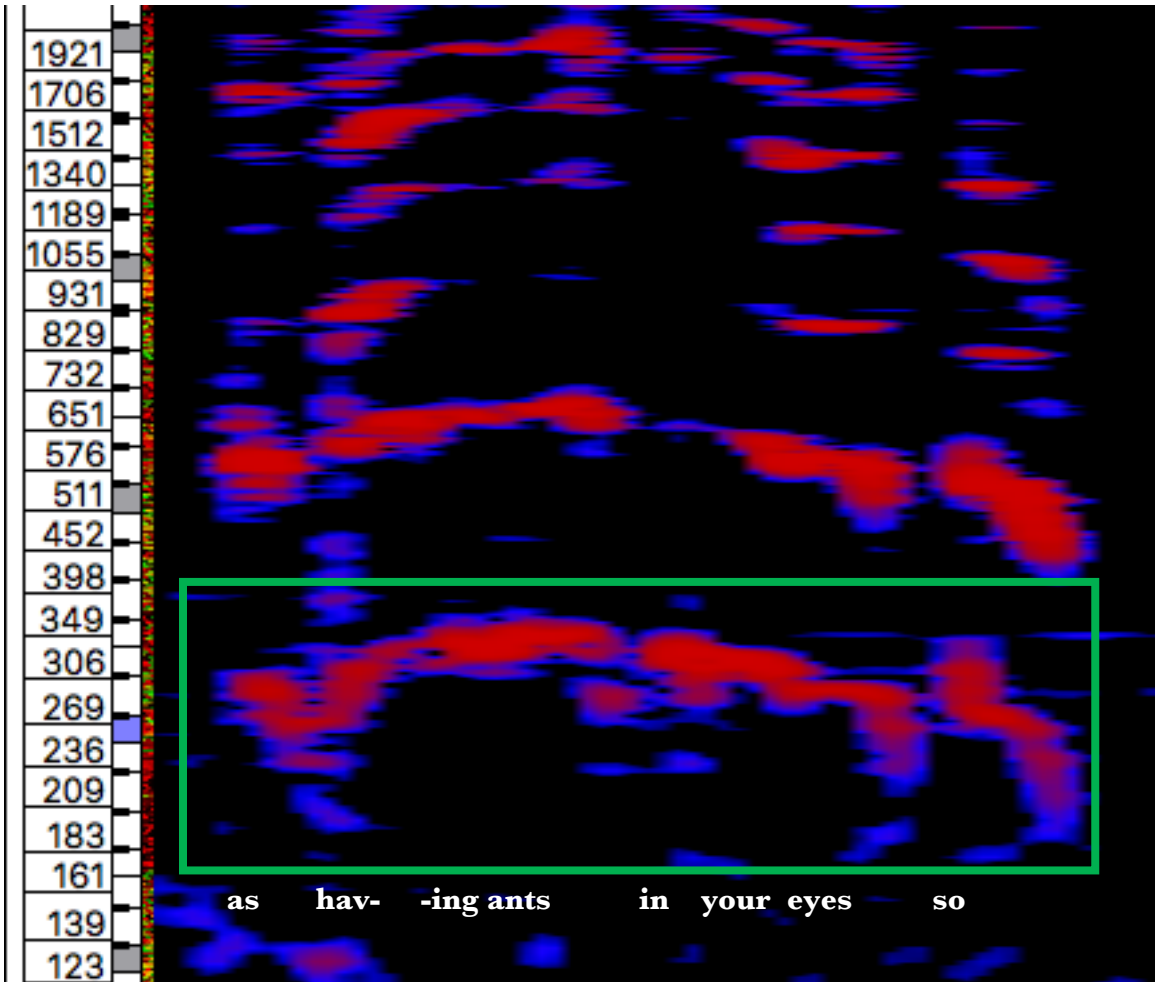
The boxes show moments of rhythmic or melodic divergence between the two lines.



Let's focus on the competing melodies that occur a bit further along in the clip which correspond to «I can't see a thing, and also I can't feel anything either. Did I mention that? But that's not as catchy as having ants in your eyes so that always goes, you know, off by the wayside» (0:27–0:37). Figure 1.7 offers a transcription of the two melodies with discrepancies between the two melodies marked. The most notable of these arises in measure 5 of the example. Where the guitar plays a diatonic melody, the sax chooses to play a chromatic line, which results in a few fleeting clashes of a minor second between the two melody instruments. The differences give the feeling that the musicians are recalling from memory a melody learned by ear, stylistically

appropriate considering the jazz-influenced style. But what's most notable here is that they're *not* drawing from a musical melody—they're interpreting the fluctuating pitch information found in the vowel formants of speech, assigning approximate fixed musical pitches to each syllable and making rhythmic decisions that likewise place the syllables on a grid of musical time.

Figure 1.8. Spectogram of «as having ants in your eyes so», Adult Swim, "Ants In My Eyes Johnson | Rick and Morty | Adult Swim," 0:36–0:37



How do these melodic lines compare to the speech itself? Figure 1.8 shows a spectrogram of the corresponding spoken phrase «as having ants in your eyes so». As we can see, the pitch contour of the phrase follows the same up-and-down gesture as both of the melodies do. As in

Figure 1.5, the first formant is marked for ease of reading. Based purely on the pitch content of the speech it seems that the guitar's step up from «having» to «ants» follows the voice more closely than the saxophone's upper neighbor gesture. On the other hand, the sax's falling chromatic motion, which corresponds to «in your eyes», more closely emulates the voice's gradual descent than does the guitar's diatonic stepwise motion. The discrepancies in this example illustrate the role of the musician in interpreting speech in a musical context, especially when "transcribing" the voice by ear. This kind of flexibility in interpretation will be a recurring theme throughout this dissertation, which includes several comparative analyses of the same speech recordings in multiple pieces.

Central questions

The various musicalizations of Ants in My Eyes Johnson's monologue that I have considered put into sharp relief many of the phenomena I wish to explore in what follows. Because its linguistic materials invite interpretation and contextualization it provides a window into possibilities for the musicalization of speech. This is especially apparent if, as I have done here, we consider successive interactions with these materials by a number of musicians, each of whom brings his/her own priorities for interpretation/contextualization through musical means. This example also begins to point toward answers to some of the research questions posed in the opening.

First, *How similar is music to speech?* Within human cultures, speech and music generally operate as different systems for communication, each with its own principles of organization, principles that may at times come into conflict with one another. Generally speaking, speech lacks stable pitch intervals while music uses fixed pitches.[10] Speech is built on a system of

[10] Aniruddh D. Patel, *Music, Language, and the Brain* (Oxford: Oxford University Press, 2008), 48.

organized timbral contrasts while music is unlikely to employ sequences of timbral contrasts for syntactic purposes.[11] Musical rhythm tends to be isochronous while speech rhythm does not.[12] Whereas natural language is built on symbolic reference, Lawrence Zbikowski has recently proposed that music is built on analogic reference.[13] To that end, as we see in the "Ants in My Eyes Johnson" example, musicians can develop strategies that put musical resources into direct contact with speech sounds by establishing and maintaining musical frameworks, or "syntactic structures that provide perceptual anchors for sonic analogs."[14] These musical frameworks set up expectations for the listener—stable pitch intervals and isochronous rhythms arranged in familiar-sounding ways. The pitch and temporal content of the speech can then be interpreted by the listener in the context of these frameworks and may either seem to conform to this musical organizational system (as in the case of "Ants in My Eyes Johnson") or not (as we'll see in later examples).

Next, *If speech and music are often associated with discrete units of auditory processing,[15] How does music adapt when confronted with spoken language as a compositional object? How can speech be heard as musical?* In "Ants in My Eyes Johnson w/Drums," we hear Dockery shifting meters and rhythmic patterns often in order to more closely mold his drumming to the rhythms of speech. That being said, Dockery's musical work isn't completely through-composed. He uses musical motives to bring out similar moments in the monologue. This happens on two levels. The first is a large-scale formal level, which can be seen in the compound-meter pattern that sets the phrase «I'm

[11] Ibid., 51.
[12] Ibid., 122.
[13] Zbikowski, *Foundations of Musical Grammar*, 167.
[14] Ibid., 128.
[15] This is assuming a domain-specific theory in which sounds are processed as either speech or music. See Diana Deutsch, Trevor Henthorn, and Rachael Lapidis, "Illusory Transformation from Speech to Song," *The Journal of the Acoustical Society of America* 129, no. 4 (April 2011): 2251.

Ants in My Eyes Johnson» (0:02–0:05 and 0:25–0:28). The second occurs on a framework-building level—the hint of metrical organization produced by the patterns that accompany «All my s—» and «all my nerves» (0:39–0:41). Changing musical frameworks to suit the rhythmic and pitch content of speech is one strategy that composers use to set speech to music, but as we'll see in later examples there are musical works in which composers purposely do not adjust their musical frameworks when the speech no long conforms within the musical system.

How do music creators—composers, songwriters, producers, and musicians—use spoken language in the production of cultural products (musical works) that do not employ language as their primary communicative resource, and how do listeners apprehend this use? First, a note about who I mean when I talk about listeners: as in much music-theoretical writing, "the listener" is to some extent really just the author—in this case, me. Nevertheless, when I discuss "listeners" I also imagine people who have heard the musical work in question and are familiar enough with it to identify its formal features. Returning to the question at hand, the answers to this will be revealed as I discuss various compositional techniques for musicalizing speech—adjustment of musical frameworks, splicing the speech recording to better suit the musical frameworks, repetition (both in the musical sense of themes and motives and in the sense of replaying snippets of speech), segmentation of speech into musical formal material, and digital manipulation of speech recordings. At this juncture, however, I can make some large-scale inferences about why these speech-driven musical cultural products are so prevalent in the last fifty or so years of Western music. I posit that audio recording is crucial to the rise of musicalized speech works—the repeatability of the medium is what allows composers and musicians the opportunity to analyze the otherwise fleeting pitch and temporal content of speech. Indeed, it is the *fixity* of recorded speech that interests me: recorded speech presents the composer with a more-or-less unchangeable body of acoustic information that has unique and essentially unmusical temporal and pitch content, and the composer must

navigate around this in the structuring of the musical work. Although I am interested in speech-as-sound, the meaning of the words being spoken—and whether or not they're understandable to listeners—will also drive my analysis of the intersecting and conflicting communicative functions of music and speech in the musical work.

I have thus far avoided one of my initial questions: *Can speech tip over into music only in selected instances, or is this kind of slippage a standard feature of human cognition?* This one example can only go so far in answering such a broad question. It seems clear following my analysis and discussion of the "Ants in My Eyes Johnson" example that Dockery and company are making a concerted effort to musicalize speech. This is certainly not the case in all instances of recorded speech accompanied by music. In fact, the question of how similarly speech and music are processed is a fraught one even in the field of cognitive psychology, which is so far split between two competing camps of speech and music perception: a domain-specific theory that assumes sounds are processed by a system that is dedicated either to speech or to music; and a cue-based theory which assumes that how a sound is processed is based on acoustic characteristics.[16] That there are so many musical works that seek to reconcile speech and music suggests that the perceptual and aesthetic line between the two is blurry.

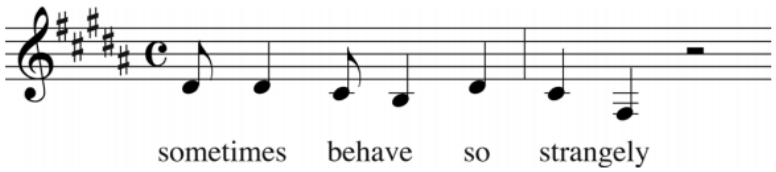
The work of psychologist Diana Deutsch suggests that, under specific circumstances, speech may be processed by the listener as music. Deutsch first demonstrated the "speech-to-song illusion," a perceptual illusion in which a spoken phrase begins to sound like song upon repeated listening, in 2003.[17] She looped a recording of the spoken phrase «sometimes behave so strangely» and found that in time a strong melodic and rhythmic profile emerged in her hearing,

---

[16] Deutsch, Henthorn, and Lapidis, "Illusory Transformation from Speech to Song," 2251.
[17] Diana Deutsch, "'But They Sometimes Behave So Strangely'," *Phantom Words and Other Curiosities*, CD (Philomel Records, 2003).

resulting in the musical phrase notated in Figure 1.9. Further study of this phenomenon by Deutsch and others supports the idea that during the process of repetition, pitches in a spoken phrase are perceptually distorted in order to fit the syntactical confines of tonal melody.[18] This of course is predicated on the assumption that a listener has either implicit or explicit expectations of the tonal system.

Figure 1.9. Musical transcription of the spoken phrase from Deutsch's "'But They Sometimes Behave So Strangely'" (reprinted from Figure 1 of Deutsch et al. 2011)[19]



It bears mention that not all speech may be processed as song by the listener. The speech-to-song illusion is rooted in artificially constructed repetitions of a selected speech segment that induce the listener to reevaluate the prosodic properties of a speech segment.[20] As it happens, however, not all speech can so readily be interpreted as song when isolated and repeated. Adam Tierney and his colleagues have studied speech passages of audiobooks, comparing samples that sound like song when isolated and repeated to samples that continue to be processed as speech

[18] Deutsch, Henthorn, and Lapidis, "Illusory Transformation from Speech to Song," 2251. See also Kankamol Jaisin et al., "The Speech-to-Song Illusion Is Reduced in Speakers of Tonal (vs. Non-Tonal) Languages," *Frontiers in Psychology* 7 (2016) and Elizabeth H. Margulis, Rhimmon Simchy-Gross, and Justin L. Black, "Pronunciation Difficulty, Temporal Regularity, and the Speech-to-Song Illusion," *Frontiers in Psychology* 6 (2015).
[19] Deutsch, Henthorn, and Lapidis, "Illusory Transformation from Speech to Song," 2246.
[20] Simone Falk, Tamara Rathcke, and Simone Dalla Bella, "When Speech Sounds Like Music," *Journal of Experimental Psychology: Human Perception and Performance* 40, no. 4 (August 2014): 1493.

under the same circumstances. They found that song-like speech contains a higher degree of fundamental frequency stability within syllables when compared with plain speech.[21]

These perceptual studies are useful in understanding some of the mechanics behind the musicalization of speech, but they do not account for all instances in which composers might musicalize speech. Composers may use recorded speech that is not readily song-like. They may avoid the repetition of spoken phrases. Whereas Deutsch's speech-to-song illusion relies solely on the sound of speech, the musical works I will discuss typically involve a more complex soundworld comprising some sort of added musical frameworks of pitch and/or rhythm that provide support for a musical hearing of speech, as happens in the "Ants in My Eyes Johnson" example. Given the complexity of the relationship between music and speech, I need to limit the scope of my project somewhat. In this dissertation, I will focus on two facets of the speech-music relationship: 1) the strategies by which composers, working within the confines of specific musical systems (e.g., tonality, a specific genre), may bring out the musicality of speech and 2) how music and natural language function in these works as means of communicating meaning.

The topic of musical works based on recorded speech brings up a large number of additional questions. What is the role of recording technology and digital audio processing in the musicalization of speech? Under what conditions do composers hear speech as musical? Does the style of the speech (poetry vs. prose) affect how it's musicalized? How are works that use musical instruments (especially those that are designed to use a musical pitch framework) different from works that use only the voice (which can be used musically but which is most often used to communicate in natural language)? What is the role or authorial status of the speaker in the

[21] Adam Tierney et al., "Speech versus Song: Multiple Pitch-Sensitive Areas Revealed by a Naturally Occurring Musical Illusion," *Cerebral Cortex* 23, no. 2 (February 2013): 251.

completed musical work? While they are not my primary research questions, the answers to these questions will be touched upon across the various chapters.

I turn now to another example, one quite different from "Ants in My Eyes Johnson," in order to explore my grounding questions about the musicalization of speech from a new angle. "Ants in My Eyes Johnson" showed us how a collection of musicians chose to add musical materials to recorded speech, materials that are in some way analogous to the sounds of speech. Conversely, American electro-acoustic composer Pamela Z's "Geekspeak" (1996) is crafted entirely out of recordings of speech with no musical instruments overlaid. The following analysis will explore how resources that are typically recruited for music—the organization of materials in time—in this case become part of how speech is presented, thus yielding a musicalization of that speech.

Pamela Z's *Geekspeak* as counterexample

Because "Geekspeak" is comprised solely of recordings of speech, it provides an excellent ground for exploring the question, *how similar is music to speech?* As this example is not supported by musical instruments, musical frameworks—to the extent that they appear—must be crafted from the sound of the voice alone. Z accomplishes this largely through the segmentation of the recorded speech into small samples and through repetition of these segments. Given that the music emerges solely from recordings of speech, the question here is not so much *How does music adapt when confronted with spoken language as a compositional object?* but instead *How can speech be heard as musical?*

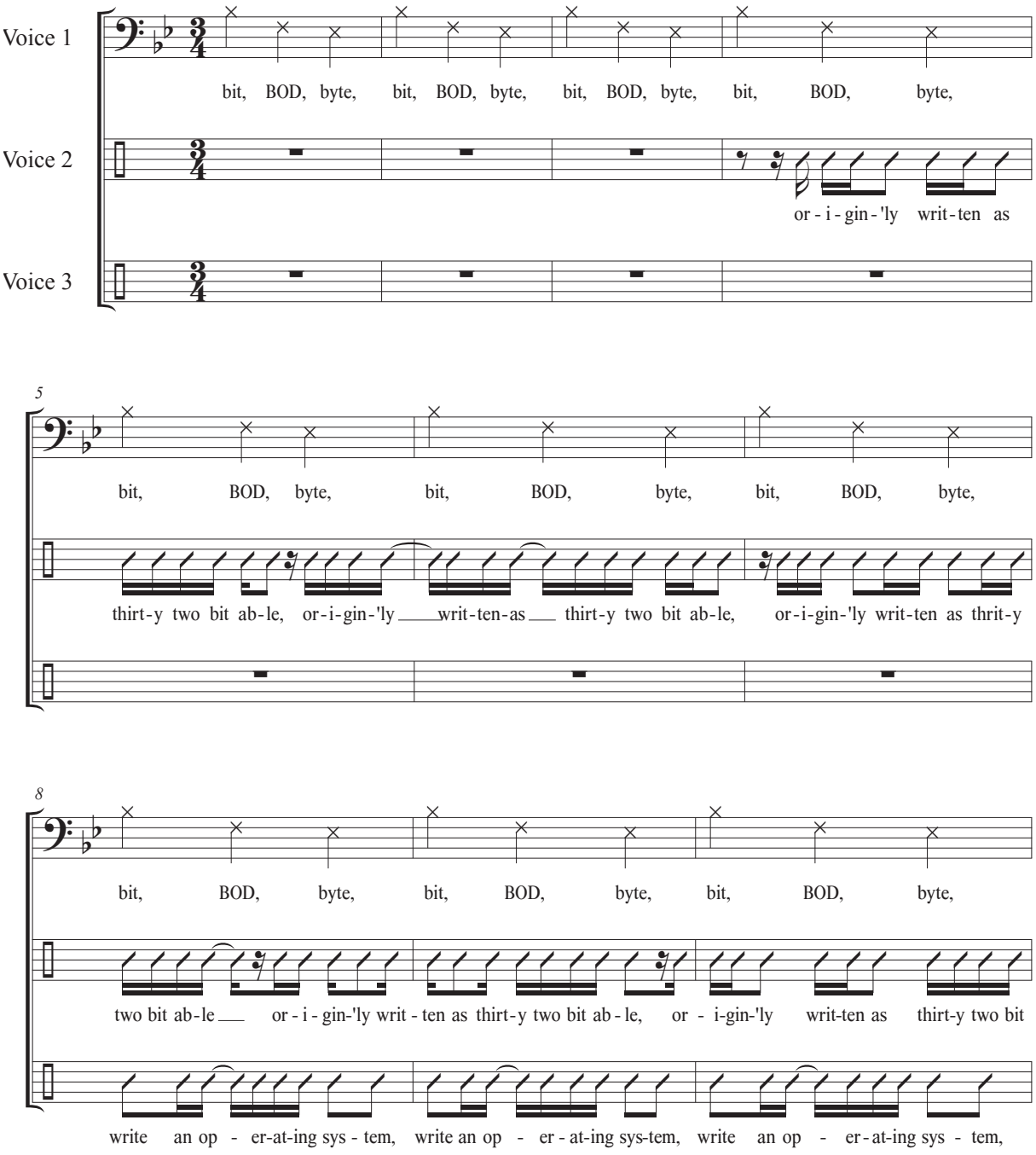
"Geekspeak" begins with the introduction of a brief sample «bit, BOD, byte», which is repeated over the first section of the piece (0:00–0:18). In the repetition of this spoken phrase, we hear the groundwork laid for a metrical hierarchy. Each syllable has roughly the same duration,

and thus provides a pulse. Because the repeating spoken phrase is three pulses (or words or syllables) in length, we hear in it the basis for a measure: a clear triple meter. Further, each syllable is successively lower in pitch, giving the highest pitched syllable «bit» a sort of accent as beat 1 of each measure.

Against this framework Z introduces two additional speech samples, one at 0:03 and the second at 0:09, both of which have much faster rates of speech. These additions put into relief the metric function of the repeated «bit, BOD, byte» segments which, in Justin London's terms, now serve as a "perceptual ground for rhythmic figures."[22] This faster rate of speech in the added speech samples results in divisions of the beat (Figure 1.10) that enrich the musical surface. Although the «origin'ly written as thirty-two bit able» motif is 4.25 beats long instead of three beats, it still works within the metrical framework due it its rhythmic subdivisions, and sets up a subtle bit of metric grouping dissonance against the metrically consonant «bit, BOD, byte» and «write an operating system». Here Z uses speech samples to create and reinforce a metrical framework through a process of repetition and layering, tipping the "unmeasured" sounds of speech over into the realm of hierarchically organized meter.

[22] Justin London, *Hearing in Time: Psychological Aspects of Musical Meter* (New York: Oxford University Press, 2004), 48.

Figure 1.10. Transcription of rhythmic motif, Pamela Z, "Geekspeak," 0:00–0:11

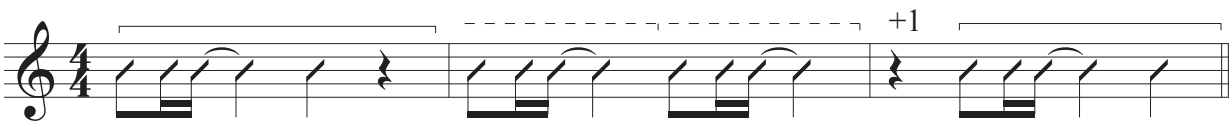


Next, *Can speech tip over into music only in selected instances, or is this kind of slippage a standard feature of human cognition?* There are moments of "Geekspeak" where I do not hear any sort of metrical framework nor any sort of particularly marked melodic material. At 0:18–0:28, Z halts

the «bit, BOD, byte» motif that had driven the metrical framework and simply presents a snippet of recorded speech from one individual. Without the metrical framework, no part of this segment stands out as being rhythmically regular from a musical standpoint. This demonstrates that some speech, particularly when presented outside of a simultaneously sounding musical context, may just not be conducive to musicalization (even though the listener has up to this point heard a metrical framework that has built up musical expectations).

But that is not to say that this recording of speech cannot be musicalized if placed into the right musical framework. Immediately following this section, Z repeats a brief sample of the speaker making a breathy paralinguistic assertion—a sound with some communicative value that yet is clearly not speech (0:29–0:35). Shown in Figure 1.11, this segment of the piece does not rely on the sort of equal-spaced repetitions seen in Figure 1.10. The four-beat motif is stated once in its complete form and then presented twice in an abbreviated form. As shown, I interpret these as two measures of 4/4 meter. Against this frame, a further iteration of the abbreviated motif enters a beat late by my expectations. The metric regularity of the two preceding measures, however, is enough—at least for me—to set up an expectation that this 4/4 meter will continue, and thus I hear the motif's placement in the third measure as a displacement dissonance. Ultimately, this snippet, which I did not hear musically in its original context, now becomes part of a musical metrical framework.

Figure 1.11. Transcription of stuttering motif, "Geekspeak," Pamela Z, 0:29–0:35



This stuttering example is particularly interesting in light of my question, *How do music creators use spoken language in the production of cultural products (musical works) that do not employ language as their primary communicative resource, and how do listeners apprehend this use?* Perhaps a more succinct question in this instance might be, *Does the semantic meaning of speech matter in these musical works?* In this example, Z is playing with a disruption in the flow of speech, a moment that does not contribute directly to our understanding of the speaker's meaning, foregrounding it and drawing it out temporally. But just because it's not significant in a linguistic sense doesn't mean it's not significant in the context of "Geekspeak." Z is calling attention to the *sound* of speech itself rather than its meaning, and highlighting the fact that not all speech need directly convey meaning. In fact, there are moments of uninterrupted, semantically complex speech within this work that are difficult for the listener to understand—for instance, I can only understand «write an» from the «write an operating system» when it returns at 2:38, not when it's first introduced over two additional, simultaneous speech phrases at 0:09. This, however, does not stop me as a listener from understanding the general topic of the speech in the work—men talking about computers and what it means to be a "geek." Just this word alone, recurring as it does throughout the work, holds a semantic charge. Nor does it stop me from making interpretive conclusions about Z's possible intended meaning for the work itself—Z is commenting on geek-dom and digital literacy by using digital technology to craft a musical work. Z is also playing with what speech can be, highlighting paralinguistic aspects of it and foregrounding the differences and similarities between music and speech. Semantic understanding is not a simple on/off situation—it may be attenuated, but the listener can still glean meaning from its expressive or paralinguistic features, or from the disconnected bits of speech that are heard semantically.

The musical structure and the compositional treatment of recorded speech in "Geekspeak" differs from "Ants in My Eyes Johnson" in two major ways. First, as previously

noted, the musical frameworks that inform our understanding of meter and tonality in "Geekspeak" are comprised solely of the speaking voice, whereas the instrumental accompaniment of "Ants in My Eyes Johnson" was a deciding factor in the listener's hearing of musical frameworks. But Z nonetheless builds musical frameworks out of pure speech—in this case by intervening in the flow of time by repeating snippets of speech. This is the basis of the second major difference: whereas in "Ants in My Eyes Johnson" the original speech recording was played through in its entirety, in "Geekspeak" Z employs repetition and technological intervention into the recorded speech object to heighten features of speech that have the potential to be used as musical resources. These two examples represent a range of methods by which composers contextually musicalize recordings of speech. This dissertation will dig deeper into these methods in order to understand how music and speech can intersect to communicate ideas and produce meaning.

Outline of the dissertation

Now that I've used two contrasting examples to introduce some of the issues surrounding the composition of music around recorded speech, I'd like to set out how these issues will play out over the course of the dissertation, which will be in two parts. The first part will deal with a range of compositional techniques for musicalizing speech with a view toward developing an understanding of the communicative resources that result. The second part will then engage with the intersections and conflicts between music and recorded speech in a range of compositions, the better to understand how music and speech convey meaning to listeners. This second part will posit the marriage of recorded speech and music as a sort of collaborative act that produces new meanings not present in the speech or music alone. Each of these larger parts will contain two chapters.

I. Compositional Techniques for Musicalizing Speech

Chapter 2 will consider some of the compositional techniques that have commonly been used to alter recorded speech to push it toward the sorts of sounds we conventionally describe as "music." I have identified these as segmentation (cutting up the recording), repetition (repeating cut-up segments), and manipulation (changing the sonic qualities of the segments). As we will see, there is a fair amount of overlap between these categories.

Recall that in Deutsch's speech-to-song illusion, it is the looping of a *small segment of speech* that gives rise to the musical qualities of speech. The curation of speech samples in musical works allows composers to choose segments of speech that lend themselves neatly to musical processes. In this section I will analyze Steve Reich's *Different Trains* from 1988, for string quartet and tape. Reich used segments of speech that are typically only one linguistic phrase in duration—translating to about one or two measures of music—to create the composition. After one iteration of the spoken voice, the musical pitch and rhythmic motives that accompany these snippets of speech are then repeated in the instruments without the voice. This "call and response" gestures bolsters the pitch and rhythmic musicality of the spoken phrase—particularly by reinforcing the spoken phrases' metrical position within the music.

The repetition of fragments of recorded speech is perhaps the most prevalent technique for musicalizing speech. Composers can splice a speech recording and loop it, as I will show Reich's tape pieces *It's Gonna Rain* (1965) and *Come Out* (1966). In these works, repetition of small segments of speech—very much like those used in Deutsch's speech-to-song illusion—defamiliarize language through challenging or even vacating its semantic content by making the listener focus on its musical features. As we saw in the Pamela Z's "Geekspeak," repetition of segments of speech can be used to provide larger structural frames for the music in terms of form

and metrical hierarchy. In this chapter, I will turn to another of Z's pieces, *Pop Titles "You"*, to demonstrate how repetition in the spoken text can be used to musical ends.

Beyond segmentation and repetition, composers can further alter the original speech recording by manipulating its pitch and temporal content on a syllable level. This technique is regularly used in the Gregory Brothers' *Auto-Tune the News* or *Songify This* YouTube series, in which speech recordings from news broadcasts, television, and movies are autotuned and played over musical accompaniment. I turn to a parody on this effect done for the theme song of the Netflix sitcom *Unbreakable Kimmy Schimdt* (2015–2019) that demonstrates how composers Jeff Richmond, together with the Gregory Brothers, used autotune to highlight the idiosyncracies of the speaker's voice in order to emphasize and musicalize speech's expressive features.

To explore a different approach to digital manipulation, I turn then to Paul Lansky's *Six Fantasies on a Poem by Thomas Campion* (1978–79). In this piece, Lansky filters and enhances a recording of his wife reading a poem by Thomas Campion in order to bring out the musicality in the pitch contours and rhythmic qualities of her speech. In each of the six movements Lansky uses speech synthesis to recreate and alter the recorded voice in several ways. I will focus on the third movement, "her reflection," which uses reverberation to "wash out" of the semantic content and prolong the pitches of vowel formant frequencies. This results in a voice that is radically altered—clearly impossible without the aid of technology—and yet still retains some markers of the speaker's humanity.

Chapter 3 focuses on compositional techniques in which the recorded speech is presented unaltered. In this approach, composers opt not to remove segments of speech that do not aurally lend themselves to musicalization, drawing on longer samples with less opportunity for repetition. This complicates the process of musicalization, but I have identified three compositional strategies that composers draw on to set these speech samples to music: repetition in the purely

musical component, adjustment of musical frameworks to suit the speech, and positioning speech as dissonant against musical frameworks.

The repetition of musical motives can be used as a means of crafting musical form. Composers can emphasize repetition of words or phrases in the speech recording by assigning them musical motives, as in the «I'm Ants in My Eyes Johnson» motif discussed earlier. In instances like this, repetition can set up the basis for the listener's hearing of musical form, making connections within the speech that might not have been apparent or significant until highlighted through the music. In this chapter I focus on "Mrs. Morris," a track from Charles Spearin's experimental jazz album *The Happiness Project* (2009). I show how Spearin uses a saxophone "doubling" of the voice to draw out repeated patterns in the pitch contours of speech, creating and developing a musical motive that unifies the track.

In some instances composers attempt to mold the musical frameworks to suit the preexisting acoustical features of speech (something evident the "Ants in My Eyes Johnson" example). James Tenney's "Song" from *Song 'n' Dance for Harry Partch* (1999) provides an example of this. In this movement, Tenney uses a 72-tone microtonal system of pitch and glissandi between pitches in order to more fully capture the nuances of vowel formant frequencies.

Composers may use recordings of speech that are not easily musicalized in order to play up the differences between speech and music, positioning speech as dissonant against musical frameworks. This approach can be heard in René Lussier's *Le trésor de la langue* (1989), an experimental rock album composed around interviews with Québec residents and speeches by historical figures about the importance of the French language is Québécois culture. The album is in a musical style typically built on a groove and a clear tonic, which are odds with long speech samples. I will examine the track "Cé ça qu'on va faire!" to show how Lussier pits recorded

speech and its musical "doubling" in the clarinet against the prevailing groove and tonic in the accompaniment.

Peter Ablinger's *Voices and Piano*, an ongoing song cycle for recorded speech and live piano begun in 1998, uses complete, largely unedited monologues as its compositional object, focusing on one speaker per movement. The movements themselves tend toward metric irregularity and atonal pitch frameworks in the piano accompaniment. Ablinger's piano accompaniment tends to place emphasis on each speaker's unique manner of speaking. As a result, the movements vary wildly in style and character, reflecting Ablinger's use of different compositional approaches to evoke the different speakers' individuality. I will use movements from this piece to show both (1) how Ablinger piece-by-piece builds a gradually emerging feeling of meter to emulate the haltingness of Arnold Schoenberg's English in the "Arnold Schoenberg" movement and (2) how Ablinger creates a musical feeling of smoothness and deepness that emulates radio announcer Bonnie Barnett's voice and yet pushes against the voice metrically.

Finally, I turn to Peter Ablinger's *Quadraturen* (1997–2004), which is a study in pushing the boundaries of music and speech. Throughout *Quadraturen*, Ablinger uses instruments to produce a kind of musical, speech-synthesis-based analysis of an unedited speaking voice. The speech is deconstructed by Fast Fourier Transform (FFT) analysis and reconstructed through a musical framework that involves discrete pitches and isochronous musical time, absent the original voice. Although played on musical instruments using "musical" frameworks, the resulting work sounds little like what we would traditionally call music—in fact, it sounds much closer to an approximation of speech. The result stands at a critical distance from true speech; that said, the

unique phonological properties of speech are still audible in the music, and, if given the text, the listener may be able to parse some of the semantic content.[23]

II. Musical Works for Recorded Speech as Collaborative Acts

This second section of the dissertation will be guided by the argument that musical works for recorded speech can be viewed as a collaboration between composers and speakers, albeit a collaboration that privileges the composer.

In Chapter 4 I will consider the role of *expression* in the musicalization of speech. Expression is a feature of both speech and music that is important in communicating moods and emotions. In particular, expression highlights the speaker as a *subject*. I argue that by using recordings of the human voice, composers are trying to engage with the humanity of the speaker, even while that humanity is radically transformed and complicated by its sonic virtual manifestation. Composers give speakers agency by positioning them as performers and by deferring some authorial control to the "texts" presented by speakers. By studying the intersection of musical and spoken expression, I highlight the communicative potential of contributions by both the composer and the speaker.

Consider again the musical phrase corresponding to «Our prices I hope aren't too low» in the "Ants in My Eyes Johnson" example (Figures 1.5–1.6). The melody of the guitar follows

[23] The resulting piece is somewhat akin to sine-wave speech, wherein three or four time-varying sinusoids replicate the estimated frequency and amplitude pattern of the resonance peaks of a natural utterance. See Robert E. Remez et al., "Speech Perception without Traditional Speech Cues," *Science* 212, no. 4497 (1981): 947–50. Ablinger's "piano speech" is necessarily different from sine-wave speech in that it uses complex tones with associated harmonic series above the sounding fundamentals—that is, the "partials" themselves are complex tones with their own spectra. The process used to create both sine-wave speech and Ablinger's piano speech, however, is essentially the same: a recording of speech is input into the computer, analyzed, and a facsimile of that speech is rendered for a single timbre based on the formant-content of the original source.

approximately the pitch information of the speaking voice to create a sort of melodic climax at the end of the phrase. This is a musical representation of a climactic, exclamatory moment in the voice. It seems likely that the expressiveness of the voice is what led the guitarist to recreate such an exclamation musically, and we should thus consider the effect of not only *speech* but *expressive speech* on the resulting music. Linguists Mark Tatham and Katherine Morton suggest that "expression is a manner of speaking, a way of externalizing feelings, attitudes, and moods—conveying information about our emotional state," defining expressive speech as "speech which gives us information, other than the plain message, about the speaker, and triggers a response in the listener."[24]

The important thing to consider in expressive speech (as well as expression in music!) is that there's more to communication than simply understanding meaning on the most basic level. In comparing speech and gesture as communicative media, for instance, David McNeill argues that "When co-expressive speech and a gesture synchronize, we see something that is both simultaneous and sequential …. There is a combination of two semiotic frameworks for the same underlying idea, each with its own expressive potential."[25] Similar processes are at work when combining speech and music, each of which come their own communicative potential. Building on McNeill's work, Lawrence Zbikowski proposes that both musical and gestural statements "interact with and shape the story that is told through language."[26] I posit, then, that combining music with recorded speech changes the way we understand the speech, adding further layers of meaning.

---

[24] Mark Tatham and Katherine Morton, *Expression in Speech* (New York: Oxford University Press, 2004), 39.
[25] David McNeill, *Gesture and Thought* (Chicago: University of Chicago Press, 2005), 91.
[26] Lawrence M. Zbikowski, "Musical Gesture and Musical Grammar: A Cognitive Approach," in *New Perspectives on Music and Gesture*, ed. Anthony Gritten and Elaine King, SEMPRE Studies in the Psychology of Music (Farnham, Surrey: Ashgate, 2011), 81.

In the repertoire of this dissertation, there is little in the way of direct one-to-one text painting, but there are, I argue, sonic analogs for dynamic processes. In Jacob TV's "Billie" (2003) the composer uses sonic analogs shaped by timbre and digital processing to highlight features of Billie Holiday's later-life vocal expression. In doing so, he focuses our attention on Holiday as an individual, placing particular focus on her ailing body at the end of her life. TV also reorders and alters speech objects and sonic space in ways that change linguistic meaning and allow musical meaning to fill space left when semantics are abandoned. By highlighting features of Holiday's vocal expression through musical means, the composer calls attention to subject position of the speaker.

As a contrasting example, I explore experimental hip hop producer DJ Spooky's 2008 remix of Pamela Z's *Pop Titles "You"*. Z's vocal delivery is quite understated, and DJ Spooky emulates the minimal structure of the spoken phrases by likewise employing repetitive, minimal musical phrases. In both this and "Billie," the recorded nature of the speech is particularly important—the effect just wouldn't be the same if the same phrases were said by a different speaker. An exploration of expressive speech will bring to the fore the role of the speaker as a human subject in the composition of these works.

Finally, Chapter 5 will consider the role of musical style in the creation of musical frameworks that serve to musicalize the voice. "Style" is a somewhat amorphous concept. Leonard Meyer defines it as "a replication of patterning, whether in human behavior or in the artifacts produced by human behavior, that results from a series of choices made within some set of constraints."[27] Although scholars have a tough time discussing style, it is one the most

[27] Leonard B. Meyer, *Style and Music: Theory, History, and Ideology* (Philadelphia: University of Pennsylvania Press, 1989), 3.

recognizable features of music for listeners.[28] I believe that the "replication of patterning" inherent to stylistic hearing forms a basis for the listener's understanding and interpretation of musical frameworks (in the repertory of this study and in music listening at large).

As such I'm interested in how recorded speech within different stylistic contexts can provide a means of *communicating* certain culturally meaningful musical signifiers to listeners. How do a composer's musical influences and personal compositional strategy affect the formation of musical frameworks? In this chapter, I propose composers employ unique musical frameworks that I call stylistic frameworks. Each musical work has a distinctive stylistic framework of pitch, time, timbre, orchestration, etc. that is decided by a variety of factors including the composer's compositional strategy, genre, and outside influences like training or personal taste.

I'll begin my musical exploration of this phenomenon by examining Charles Dodge's *The Story of Our Lives* (1973–74), a computer synthesis piece that uses speech analysis and synthesis to convert plain speech into a recitative-like vocal delivery. Recitative itself has long been held to be a sort of heightened artistic interpretation of speech sounds. I will explore which features of recitative as a musical genre Dodge uses to musicalizes genuine speech and how Dodge sets his singing speech synthesis apart from speech itself. In doing so, I explore the ways in which generic and individual composer's stylistic frameworks are used to musicalize recorded speech.

I turn then to Peter Ablinger's *Voices and Piano*, a piece in which the musical frameworks are influenced not only by genre, but also by the *compositional style* of the speaker whose voice is heard in the work. In particular, I will explore the "Arnold Schoenberg" movement, in which the music Ablinger writes to accompany Schoenberg's recorded voice recalls, through Ablinger's

[28] See Robert O. Gjerdingen and David Perrott, "Scanning the Dial: The Rapid Recognition of Music Genres," *Journal of New Music Research* 37, no. 2 (2008): 93–100; Carol L. Krumhansl, "Plink: 'Thin Slices" of Music,' *Music Perception* 27, no. 5 (2010): 337–354.

organization of rhythms, textures, and dynamics, Schoenberg's compositional style. This compositional stylistic allusion conveys referential meaning by evoking extra-musical association with Schoenberg's music and hence Ablinger is able to interact with Schoenberg as a composer, an historical figure, and, finally, a human subject with a voice.[29]

Speech and the voice

The reader may notice that I'm referring to "speech" or "the speaking voice" instead of simply using the term "the voice." This is because the voice is a rather broad topic in contemporary music scholarship and I feel that I need to be clear about what aspect of the voice interests me in this dissertation: the voice as actuality, as sound, as a carrier of natural language.

The voice as metaphor has of course proliferated in music academic discourse. Perhaps the best illustration of this comes in Martha Feldman's introduction to the 2015 *JAMS* colloquy "Why Voice Now?" Feldman offers an overview of the many different ways music academics deploy "the voice": physically, psychoanalytically, linguistically, sonically, embodied or disembodied.[30] Edward Cone's *The Composer's Voice* (1974) famously theorized the voice in music as an extension of the composer's persona.[31] We often talk of instrumental "voices" in analysis, as Fred Maus has explored in account of agency in instrumental music and song.[32] Carolyn Abbate imagined a narrative voice in opera and envisioned the physical voices from human bodies as the

[29] This definition of stylistic allusion is borrowed from Peter Manuel, "Music as Symbol, Music as Simulacrum: Postmodern, Pre-Modern, and Modern Aesthetics in Subcultural Popular Musics," *Popular Music* 14, no. 2 (1995): 231.
[30] Martha Feldman, "The Interstitial Voice: An Opening," *Journal of the American Musicological Society* 68, no. 3 (2015): 653–59.
[31] Edward T. Cone, *The Composer's Voice* (Berkeley: University of California Press, 1974).
[32] Fred Everett Maus, "Agency in Instrumental Music and Song," *The College Music Symposium* 29 (1989): 42–43.

means of production of these prosopopoeic voices.[33] There are myriad ways of talking about the voice in music, but I wish to limit the parameters in order to explore the very specific phenomenon of using recorded speaking voices in composition.

The voice can be used in a traditional musical sense (singing) while also carrying linguistic information, but the delivery of its linguistic content is invariably reshaped or even distorted when words become song. Lawrence Zbikowski explores this phenomenon at length in analyses of several musical settings of Goethe's "Über allen Gipfeln ist Ruh," and concludes that setting language to song has the effect of "altering the pronunciation of words, slowing their delivery, introducing repetitions and pauses, and constraining the images to which they give rise."[34] While the composers in my study may introduce repetitions and pauses into their musical settings by manipulating the recording, the pronunciation of the words and their delivery remain unchanged. This is important because, within these compositions, one still hears the expressive content of the speech while simultaneously hearing the expressive content of music.

The voice is not the only way to deliver language—consider sign language and writing—but is the only means rendered in sound (and I would maintain this is so even in the case of synthesized speech). This study is concerned with the phonological content of the speech object—how *speech sounds* become *musical sounds*. But beyond the voice as a sound source, I am interested in *language* as it is delivered through the voice and its tensions with the *paralinguistic aspects of music*. While the phonological content of speech is interesting to me on a sonic level, the semantic content is also important as it can inform our hearing of speech's expressiveness. I will

[33] Carolyn Abbate, *Unsung Voices: Opera and Musical Narrative in the Nineteenth Century* (Princeton, N.J.: Princeton University Press, 1991).
[34] Zbikowski, *Foundations of Musical Grammar*, 198. See Chapter 6, "Music and Words," 167–200.

therefore be engaging with how speech's semantic meaning may contribute to a listener's experience of musicalized speech.

The composers whose work I discuss in this dissertation do not treat the semantic content of their speech objects in any sort of unified way. As I have already noted, speech's meaning can be foregrounded by repetition or its meaning can be pushed aside by segmentation. But what listeners take away from speech can vary: in some cases they may take different meanings away from the linguistic content of different performances of the same piece. I have almost no trouble understanding what the voice is saying in pianist Mark Knoop's recording of Ablinger's *Voices and Piano*.[35] At a 2018 live performance of this work by pianist Daniel Pesca, however, I found it nearly impossible to make out the voice beyond simply hearing that there was a voice speaking. In both cases, the amplitude level of the recorded speech relative to the piano reflected a conscious decision on the part of the performer. All this to say, understanding the semantic content of a musicalized speech work is highly listener-subjective, but the meaning of the speech should not be written off because of this—it contributes to our interpretation of the musical work and how we engage with both music and speech within the work as a means of communication.

This study is hardly the first to consider speech in relationship to music. As but one example, during the French Enlightenment Rousseau and Condillac envisioned music as an originary language.[36] Czech composer Leoš Janáček was obsessed with speech melody in the late 19th and early 20th centuries, believing that an understanding of the rhythmic and melodic

---

[35] Mark Knoop, *Popular Contexts & Voices and Piano*, CD, vol. SR382 (Brussels: Sub Rosa, 2013).
[36] See Jean-Jacques Rousseau, "Essay on the Origin of Languages (c. 1763)," in *Essay on the Origin of Languages and Writings Related to Music*, trans. John T. Scott, vol. 7, The Collected Writings of Rousseau (Hanover: University Press of New England, 1998 [1763]), 289–332; Etienne Bonnot de Condillac, *An Essay on the Origin of Human Knowledge; Being a Supplement to Mr. Locke's Essay on the Human Understanding.*, trans. Thomas Nugent, History of Psychology Series (Gainesville, Fla.: Scholars' Facsimiles & Reprints, 1971 [1756]).

contours of speech is essential to operatic vocal writing.[37] Arnold Schoenberg famously used *Sprechstimme* in his *Pierrot Lunaire* (1912), asking the performer to adhere to notated musical rhythms but to avoid the discrete, held pitches associated with music in order to produce a vocal technique that falls between music and speech.[38] By mid-century, *musique concrète* composer Pierre Schaeffer mused on the relationship between music and language, noting that while the two are largely dissimilar the rules of language have analogs in music.[39] In 1963, ethnomusicologist George List theorized intermediate forms between speech and song (recitation, intonational recitation, chant, and intonational chant) using analyses of recordings of the voice created using a spectrograph.[40] In more recent scholarship, Lawrence Kramer explored the idea of "melodic speech," a speech in which the speech melody (always present because all speech has fluctuating pitch content) is marked to the listener.[41] Steven Rings has considered speech and speech-song in performers of many popular music styles in the late 20th century, with specific focus on country

[37] Paul Wingfield, "Janáček's Speech-Melody Theory in Concept and Practice," *Cambridge Opera Journal* 4, no. 3 (1992): 283–84. See also John Tyrrell, "Janáček and the Speech-Melody Myth," *The Musical Times* 111, no. 1530 (1970): 793–96; Paul Christiansen, "The Meaning of Speech Melody for Leos Janáček," *Journal of Musicological Research* 23 (2003): 241–63.
[38] Arnold Schoenberg, *Verklärte Nacht and Pierrot Lunaire* (New York: Dover Publications, 1994), 54. See also Julia Merrill and Pauline Larrouy-Maestri, "Vocal Features of Song and Speech: Insights from Schoenberg's Pierrot Lunaire," *Frontiers in Psychology* 8 (2017): 1108, which explores which features of speech and song listeners attend to while hearing performances of *Sprechstimme*.
[39] Pierre Schaeffer, *Treatise on Musical Objects: An Essay across Disciplines*, trans. Christine North and John Dack, vol. 20, California Studies in 20th-Century Music (Oakland, California: University of California Press, 2017), 205–248. Schaeffer compares Roman Jakobson's two modes of arrangement (selection—choosing between words with similar meanings—and combination—parsing linguistic units into smaller or larger linguistic entities) to timbral/orchestration choices and melody's parsability into both individual notes and larger formal structures (234–37).
[40] George List, "The Boundaries of Speech and Song," *Ethnomusicology* 7, no. 1 (1963): 1–16.
[41] Lawrence Kramer, *Expression and Truth: On the Music of Knowledge* (Berkeley: University of California Press, 2012), 72–87.

singer-songwriter Hank Williams, R&B singer Marvin Gaye, proto-punk Patti Smith, and folk musician Laura Marling, Chance the Rapper, and Beyoncé.[42]

This is also not the first study to consider the musicalization of recorded speech. G. Douglas Barrett has written on the intersection of music, language, and noise in Peter Ablinger's *A Letter from Schoenberg* (2007) through the lens of Deleuze and Guttari's "perceptual semiotics."[43] James O'Callaghan has surveyed works in the spectral music tradition that use what he calls "mimetic instrumental synthesis" to recreate complex recorded sounds (including speech and birdcalls) through instrumental musical means using spectral analysis.[44] Several scholars who have considered recorded speech in music have examined the subjectivity of the speaker. Steven Connor views the recorded voice in postmodern music as decomposed and degraded, incapable of recapturing its former humanity.[45] Maarten Beirens sees violence in such degradation of the voice in Steve Reich's tape works for voice, *It's Gonna Rain* and *Come Out*, which repeat speech fragments of African American speakers until they stripped of their semantic elements and become purely sonic objects.[46] Luke Howard views the disembodied recorded voice in works like Reich's as "Godly speech," projecting a sense of distant, unquestioned authority.[47] David Code

---

[42] Steven Rings, "Speech and/in Song," in *The Voice as Something More: Essays toward Materiality*, ed. Martha Feldman and Judith T. Zeitlin (Chicago: University of Chicago Press, 2019), 37–53.
[43] G. Douglas Barrett, *After Sound: Toward a Critical Music* (New York: Bloomsbury Academic, 2016), 96–115.
[44] James O'Callaghan, "Mimetic Instrumental Resynthesis," *Organised Sound* 20, no. 2 (August 2015): 231–40.
[45] Steven Connor, "The Decomposing Voice of Postmodern Music," *New Literary History* 32, no. 3 (2001): 479–80.
[46] Maarten Beirens, "Voices, Violence and Meaning: Transformations of Speech Samples in Works by David Byrne, Brian Eno and Steve Reich," *Contemporary Music Review* 33, no. 2 (2014): 217.
[47] Luke Howard, "The Voice(-Over) of God: Some Thoughts on the Disembodied Voice in Contemporary Music," *The Open Space Magazine* 1 (1999): 109–116.

argues for a loss of identity that the speaker undergoes in Paul Lansky's *Six Fantasies on a Poem by Thomas Campion* as the intelligibility of the speech is lost in the first, third, and fifth movements.[48]

This scholarship notwithstanding, what I offer here is, to the best of my knowledge, the first study of the musicalization of speech from an in-depth music analytical perspective. I mean to explore how composers structure their music around the sounds of speech, which at times are seemingly incompatible with systems of musical grammar. I also want to examine the resulting *sonic* product is heard and understood by listeners. At the heart of this dissertation, I am concerned with how speech and music convey meaning to listeners—and how they interact or conflict in doing so when put together.

[48] David Loberg Code, "Observations in the Art of Speech: Paul Lansky's Six Fantasies," *Perspectives of New Music* 28, no. 1 (1990): 167.

Part I: Compositional Techniques for Musicalizing Speech

Chapter 2: The Alteration of Speech

In this first part of the dissertation, I will examine compositional techniques that composers use to musicalize speech—in effect, to slot the sounds of speech into musical frameworks. How to suit words to music is an age-old problem, one that is only amplified by the present repertory. As Eric Prieto observes,

> At the most elementary level, [a composer] must decide whether (to paraphrase Monteverdi) the music should be the servant or the master of the words. Should the acoustic properties defined in the score be the primary focus of the listener's attention (with the lyrics serving in a support capacity, i.e., as an incitement to invention on the part of the composer or as a guide to interpretation for the listener) or should the music play the supporting role (as, say, an intensifier or expressive supplement)?[1]

Prieto goes on to explore the influence of the text—particularly recorded speech as text—in Steve Reich's multimedia opera *The Cave* (1993), in which small snippets of speech recordings for the rhythmic, melodic, and harmonic basis for much of the music:

> This approach to the problem of text setting is of interest to contemporary composers for several reasons: for the innovative way in which it integrates digital technology into the compositional process; for its tendency to steer melodic invention into new channels, which remain intelligible to the casual listener while at the same time providing interesting intellectual challenges to the dedicated listener; and—most significantly—for the way in which it modifies the relationship between music and text, not only shifting the balance of power between them, but fundamentally altering the ways in which musical structure and semantic content work together.[2]

These issues are in play not just in *The Cave*, or in Reich's other works for recorded speech—including *It's Gonna Rain* (1965), *Come Out* (1966), *Different Trains* (1988), and *WTC 9/11* (2011)—but across the broader repertory of musical works involving recorded speech. Composers in this oeuvre may make the music the master of the words by altering the speech recordings in some

[1] Eric Prieto, "Speech Melody and the Evolution of the Minimalist Aesthetic in Steve Reich's *The Cave*," *Circuit: Musiques Contemporaines* 12, no. 2 (2002): 21.
[2] Ibid., 22.

way, or make the music the servant by presenting the speech recordings largely unedited and drawing musical materials from them. While composers have long dealt with setting text to music, this particular process, as Prieto suggests, is unique to the modern era in its application of technological mediation, using speech, words, and natural language in a way that composers who worked before the advent and widespread use of this technology could not.[3] My intent in the first part of this dissertation is to survey the ways in which music interacts with recorded speech, taking inventory of common compositional techniques and strategies that are present in this repertory.

My focus in the current chapter will be on compositional techniques that alter the speech in some way in order to suit the musical frameworks being employed. The following chapter will then turn to the compositional strategies deployed in musical works in which the recorded speech is presented in a relatively unaltered form. The three main categories of altering recorded speech upon which I shall focus are (1) the act of segmentation, or splicing recorded speech into smaller fragments; (2) repetition, the recurrence of specific events (e.g., tape loops as direct repetition or the repetition of words or textual themes as indirect repetition); and (3) technological manipulation, by which composers use recording and sound processing technologies to change the pitch, rhythmic, or timbral qualities of speech to better suit the musical frameworks. I have ordered them here by the level of intervention required on the part of the composers: there's first segmentation, cutting up the recording; then repetition, repeating cut-up segments; and then manipulation, changing the sonic qualities of the segments. My main concern in the following

[3] There are certainly many interesting examples of composers trying to capture the qualities of speech through music without the help of recording technology—recitative in operatic works, Arnold Schoenberg's *Sprechstimme*, Leos Janacek's theory of speech melody. These are of course quite interesting, but my interest lies in works for recorded speech because composers often use these recordings to incorporate *very specific, analyzable, repeatable* rhythmic and melodic details from speech into their music.

analyses is to address two connected questions: What does the music add to these recordings of speech, and how does setting these recordings to music affect the expression of ideas, thoughts, and emotions?

One of the reasons that the fixity and repeatability of the recorded medium is so important to these compositions is that each utterance of speech is unique, and recordings capture a particular performance. I find this quote from linguist Edmund Gussman particularly pertinent:

> Strictly speaking, no two sounds are ever exactly identical even if they are perceived as such by users of the language: there are individual differences between speakers as far as their voice quality goes, and even the same speaker on different occasions will produce sounds that differ, for example, in loudness… All linguistic practice tends to disregard such minute phonetic distinctions, but this means the sounds we speak of are in reality not physical but *abstract sounds*.[4]

Gussman thus urges caution when approaching speech sounds in linguistic analysis, lest those sounds be viewed monolithically. Hence, in my analyses I do wish to talk about *actual* speech sounds—it is these differences in voice quality and delivery that the composers in this repertory have latched onto and sought to heighten by way of setting them to music.

What about the comprehensibility of the speech? Ultimately, these works seem to be about playing with the *sound* of speech, its phonetic and phonological features rather than its semantic ones. That said, all of these examples present the recorded speech in the form of statements that are intelligible to the listener (at least in some of their iterations over the course of the piece).[5] Thus I believe that the meaning of the speech plays a somewhat foundational role in the composers' structuring of the musical work and informs the listeners' interpretation of that

---

[4] Edmund Gussmann, *Phonology: Analysis and Theory* (Cambridge, U.K.: Cambridge University Press, 2002), 3. Emphasis in original.

[5] A counterexample is provided by Paul Lansky's *Idle Chatter* (1985), in which recordings of Hannah Mackay's speech are fragmented into syllables and sub-syllabic elements, and reordered. The result is recognizably derived from language, but can't be understood semantically.

work, even if in some works the speech may have been chosen for its acoustical features more so than its meaning. As such, I will explore the text-music relationship in each analysis, gesturing toward how the text (in addition to the speech as sounded) informs both the musical structure and the listener's apprehension of the whole.

The following analyses are by no means meant to be a how-to of composing for recorded speech. My goal in analyzing these works is to survey the compositional techniques that composers have used—and how they're employed across genres and art movements—in the hopes of getting to the bottom of why so many composers and musicians across so many styles of music in the modern era have experimented with this type of composition. I hypothesize that it is the result of a confluence of new technologies for sound production/reproduction and the rise of experimentation in incorporating everyday, traditionally non-musical sounds into musical works—speech sounds in particular because they bring with them a pre-existing system of organization that intersects and conflicts with musical organization in interesting ways..

2.1. Segmentation

Segmentation is an important concept in both language and music. According to Raymond Monelle, segments are simply "those shorter or longer bits of language or music."[6] These can be examined and differentiated from two points of view. First, one can explore how they are strung together syntagmatically, with the focus on how a word or sound relates to other words or sounds. Second, one can group them together paradigmatically, concerning how a word or sound is differentiated from others.[7] Monelle goes on to contrast how syntagmatic and

[6] Raymond Monelle, *Linguistics and Semiotics in Music* (Chur, Switzerland: Harwood Academic, 1992), 59.
[7] Ibid. Monelle borrows the idea of syntagm and paradigm from Saussure; see also pages 32–34 for Monelle's introduction to Saussure's terminology. Nicolas Ruwet, *Langage, musique, poésie*, Collection Poétique (Paris: Éditions du Seuil, 1972) argues for a similar position on musical

paradigmatic organization function in music and language. Repetition is important to the paradigmatic organization of both language and music (e.g., a recurring phoneme in a conversation or a recurring motive that unifies a long passage of music). Syntagmatic repetition, however, is more common in music. It's unlikely that a phoneme will repeat successively, but the direct repetition of a 2-bar phrase in a lyric melody is an oft-used musical gesture.[8] This difference gets at the heart of how the carefully curated segmentation and repetition of speech inform its switch from the linguistic to the musical domain in these musical works: by choosing segments of speech, divorcing them from their original linguistic context and repeating them, composers can shift the focus from the meaning of speech to the sound of speech, highlighting the more "musical" features like melodic contour and rhythm.

For her part, Dora Hanninen defines a segment in music as a grouping of notes (or sound-events) that one takes to be a significant musical object in an analytic discourse.[9] She theorizes segmentation as an act of contextualization based on a listener's hearing of association and disjunction. Association is supported by repetition, equivalence, or similarity and results in cognitive chunking at the level of individual segments and at higher levels of organization.[10] Disjunction identifies perceptual salience with difference, which locates boundaries that imply groupings.[11] While Hanninen's theory is designed for music, the idea that listeners cognitively

---

segmentation, wherein analysis yields categories of like segments (akin to Monelle's paradigmatic segmentation) and a hierarchy of their combinations (akin to Monelle's syntagmatic segmentation). See also David Lidov, *Is Language a Music?: Writings on Musical Form and Signification* (Bloomington: Indiana University Press, 2005), 26–27 for an English-language gloss on Ruwet.

[8] Ibid., 65–66.

[9] Dora A. Hanninen, *A Theory of Music Analysis: On Segmentation and Associative Organization* (Rochester, NY: University of Rochester Press, 2012), 63.

[10] Ibid., 20.

[11] Ibid., 19.

and analytically construct segments to aid in comprehension applies to both music and speech.[12] Both Monelle and Hanninen are concerned with how *listeners* perceive difference and similarity between sounds in time, and the differences between the organization of linguistic and musical sound provide interesting opportunities for interplay and conflict between these two mediums of communication.

The creation of a musical work from recorded speech requires some level of "musical analysis" on the part of the composer, who first must hear and identify a potential for musical sound in speech before reconfiguring speech recordings in a way that brings out this musicality. In order to demonstrate how segmentation of speech informs musical structures in such works, I turn now to a composition by Reich: *Different Trains* (1988), which is scored for tape and string quartet. The piece was inspired by Reich's reminiscences of riding trains cross-country as a child to travel between his mother in Los Angeles and his father in New York between 1939 and 1942, and his reflection on the trains that would have transported him to internment camps had he, as a Jew, lived in Europe during this time.[13] The piece includes samples of recordings that relate to trains in Europe and the United States before, during, and after World War II. The first movement, "America—Before the War" features samples from recorded audio interviews with two figures related to Reich's own travels: Reich's childhood governess, Virginia, who traveled with him on these train rides; and Lawrence Davis, a retired Pullman porter who worked on the lines between New York and Los Angeles during this time. The second movement, "Europe—During the War" is built around recordings of the voices of Holocaust survivors Rachella, Paul,

---

[12] For more on how listeners segment speech, see Anne Cutler, Delphin Dahan, and Wilma Van Donselaar, "Prosody in the Comprehension of Spoken Language: A Literature Review," *Language and Speech* 40, no. 2 (1997): 159–71, which focuses on the computation of syntactic structure.
[13] For more on Reich's conception of the piece, see Reich's liner notes for Kronos Quartet and Pat Metheny, *Different Trains / Electric Counterpoint*, CD, 79176-2 (New York: Elektra Nonesuch, 1989).

and Rachel, drawn from the Fortunoff Video Archive for Holocaust Testimonies, Yale University Library, and the Holocaust Collection of the American Jewish Committee's William E. Weiner Oral History Library. The third movement, "After the War" intermixes the voices of both the American and European speakers. The melodic and rhythmic content of the instrumental string quartet accompaniment is informed by the sounds of the recorded speech. This orchestrational context brings up a question on which I would like to focus in the following analysis: how does the inclusion of traditionally musical instrumental sounds affect the way in which the composer segments speech into musical form?

The answer provided by Reich's *Different Trains* is that each spoken phrase gets its own musical section with a distinct rhythmic/melodic motif, meter, tempo, and key. To demonstrate this, let me turn to the first movement, "America—Before the War." (See Figure 2.1 for a diagram of the form of this movement.) One speech sample—one spoken phrase—is repeated within each section, emphasizing the musical rhythmic and melodic qualities that Reich hears in these recorded speech samples. But the spoken phrase does not account for all of the musical material. In each section, the main motif, played by one of the four instruments of the ensemble, is accompanied by a sort of chugging, sixteenth-note accompaniment that serves as a sonic analogue for the steady, driving motion of a train. Although the speech recordings are typically incomplete phrases, the overarching theme of train travel across the United States in the years preceding World War II emerges through keywords (city names, years, the word «trains»).

Figure 2.1. Formal diagram, Steve Reich, *Different Trains*, I. "America Before the War"

Motif	Train whistle	«from Chicago to New York»	«one of the fastest trains»	«the crack train from New York»	«from New York to Los Angeles»	«different trains every time»
Speaker		Virginia		Mr. Davis		Virginia
Timing	0:00–0:35	0:35–1:22	1:22–2:36	2:36–3:42	3:42–4:41	4:41–5:42
Notated key	F	D♭	A♭	C	D♭	G♭
Tempo	♩ = 94.2	♩ =108	♩ = 97	♩ = 84	♩ = 69	♩ = 76
Meter	2/4	2/4 + 3/8	2/4	2/4	2/4	2/4

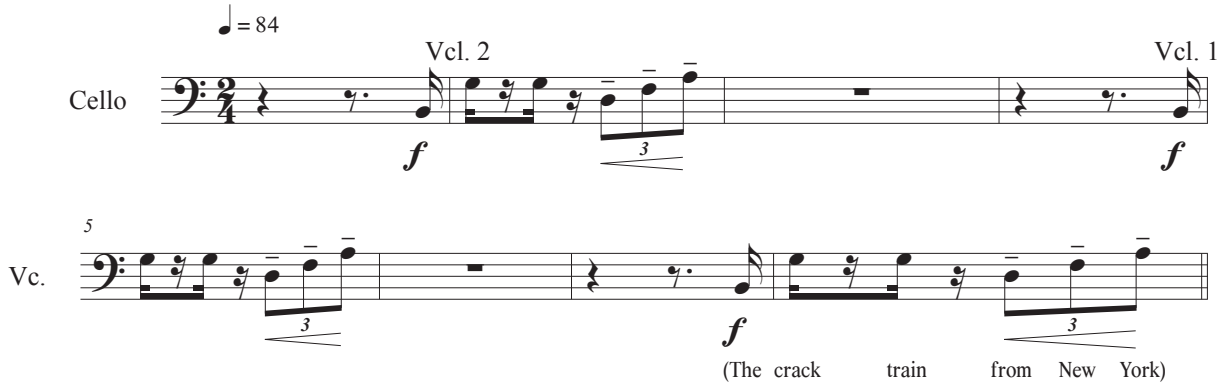
Motif	«from Chicago to New York»	«in nineteen thirty-nine»	«nineteen thirty-nine»	«nineteen forty»	«nineteen forty-one»	«nineteen forty-one I guess it was»
Speaker	Virginia		Mr. Davis			Virginia
Timing	5:42–6:21	6:21–7:01	7:01–7:35	7:35–8:03	8:03–8:29	8:29–8:59
Notated key	D♭	E♭	E	E	A♭	E♭
Tempo	♩ = 108	♩ = 130	♩ = 126	♩ = 126	♩ = 126	♩ = 99
Meter	2/4 + 3/8	3/4	5/4	5/4	3/4	3/4

The melodic/rhythmic motifs are often set up by the instruments first before being coupled with the recorded speech from which their temporal and pitch content is derived. Take for instance the fourth section of the movement (2:36–3:42).[14] The measure-length motif is heard first in the pre-recorded cello 2 part at 2:37, and then three measures later in the live cello 1 at 2:41 (see Figure 2.2 for the notation for these passages). When the voice enters with the phrase «the crack train from New York»—once again three measures after the onset of the previous iteration of the motive (at 2:46)—we've heard Reich's melodic/rhythmic interpretation of this

[14] Timings taken from Kronos Quartet and Pat Metheny, *Different Trains / Electric Counterpoint*, CD, 79176-2 (New York: Elektra Nonesuch, 1989). Because the piece uses a pre-recorded tape, other recordings are likely to have similar timings.

spoken phrase twice already in quick succession, and are primed to hear the pitches and rhythms of the speech within this given musical framework. The repetition has also primed us to hear the motive as periodic, part of a three-measure cycle. Ultimately, the segment of the speech recording (the phrase «the crack train to New York») is what forms the rhythmic and melodic basis for the measure-length musical motif. The three-measure cyclic musical phrase appears to be a purely musical decision, one that elongates the phrase and shifts the focus from the voice/melody to other musical parameters, namely the dense texture and locomotive rhythm of the accompaniment.

Figure 2.2. Cellos and tape, Steve Reich, *Different Trains*, "I. America Before the War," mm. 137–144, 2:36–2:48

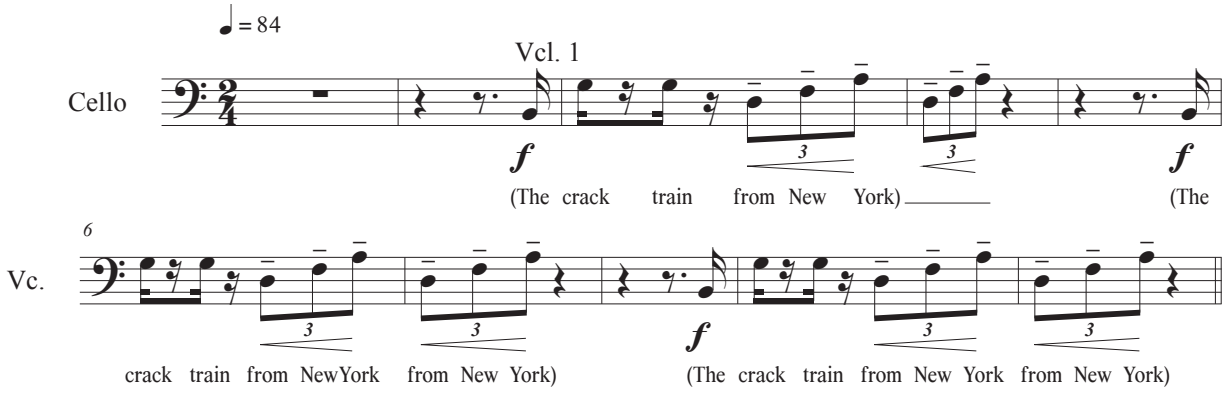


Immediately following the excerpt shown in Figure 2.2, Reich uses repetition and segmentation on a smaller scale to expand the motif. In this case, Reich repeats the triplet arpeggio that underscores the last three syllables of the phrase (a phrase that's still understandable on its own—the prepositional phrase «to New York») and shortens the rest following this repetition in order to maintain the three-measure musical phrase (Figure 2.3).[15]

[15] Immediately following this, the cycle is shortened to two measures, providing a sort of hypermetrical acceleration.

Again, this expansion of the motif is introduced melodically and rhythmically by the cello alone before the speech is revealed. By introducing the melody before the speech sample, Reich is foregrounding the pitch and temporal qualities of speech. By the time the speech recording is heard, the listener already has an established musical framework through which to interpret the phonological features of the spoken words.

Figure 2.3. Cellos and tape, Steve Reich, *Different Trains*, "I. America Before the War," mm. 145–154, 2:47–3:01



Ultimately, this example demonstrates how segmentation may take place on different levels of the structural hierarchy of both language and music in a way that informs our hearing of this piece as *musical*. It furthermore shows how segmentation and repetition can musicalize speech by making the listener consider it syntagmatically, outside the realm of everyday speech communication. Reich has chosen to focus the listener's attention on small segments of speech, abstracted from their original context, which are emphatically repeated in small phrases over large sections of the music. This approach to formal organization is a hallmark of the minimalist style in which Reich is writing, a style that several scholars see as an attempt to mute or reconfigure musical expression. Guillaume Pastre asserts that minimalism as a style does not value expression, placing more aesthetic emphasis on repetition, the organization of materials,

and the accessibility of the structure.[16] According to Wim Mertens, minimal music is "is non-representational and is no longer a medium for the expression of subjective feelings."[17] Steven Connor and Maarten Beirens have similarly argued that the disembodied human voice in minimal music degrades the voice and strips it of its humanity, using Steve Reich's music to bolster their claim.[18] I believe that we can hear this in *Different Trains*. By repeating the same spoken phrase over and over again, Reich de-emphasizes the semantic aspect of the phrase in favor of the phonetic or phonological, akin to his earlier tapes works. But this treatment does not suppress our ability as listeners to identify speech as speech. Speech is sound, yes, but a special type of sound made by humans and intelligible enough to retain its symbolic features. Even though Reich may alter or bury some of its intelligibility through repetition and the layering of musical instruments, we can still identify it as speech—a type of sound that carries different meanings and structures than music. In this analysis that follows, I consider the place of linguistic meaning in the music.

2.2. Repetition

In *On Repeat* Elizabeth Margulis begins with a comparison of music and language as means of communication, identifying several cultural attitudes that have contributed to the primacy of language over music in cognitive science research: (1) that the production of music requires a level of specialization not present in language communication; (2) that music is

[16] Guillaume Pastre, *Un art de la cohérence: Different Trains, Steve Reich*, Univers musical (Paris: L'Harmattan, 2018), 36.
[17] Wim Mertens, *American Minimal Music: La Monte Young, Terry Riley, Steve Reich, Philip Glass*, trans. J. Hautekiet (London: Kahn & Averill, 1983), 88.
[18] Steven Connor, "The Decomposing Voice of Postmodern Music," *New Literary History* 32, no. 3 (2001): 467–483; Maarten Beirens, "Voices, Violence and Meaning: Transformations of Speech Samples in Works by David Byrne, Brian Eno and Steve Reich," *Contemporary Music Review* 33, no. 2 (2014): 210–222.

ineffable and experienced subjectively; (3) that language is essential to human identity while music is an excess meant to excite our pleasure circuitry; and (4) that repetition is ubiquitous in music, while there is a "tendency to view repetition as regressive, childlike, and embarrassing" when it occurs in language.[19] But Margulis goes on to defend the exploration of repetition's role in music: "Repetition is not an arbitrary characteristic that has arisen in a particular style of music; rather, it is a fundamental characteristic of what we experience as music."[20] Margulis's comparison of how music and language are treated in Western thought has particular resonances with this study: music and language are (often justifiably) considered to be vastly different means of communication, and repetition is typically associated more with the domain of music.[21] Repetition then is a particularly powerful means of pushing speech sounds toward being heard as musical.

Recall from Chapter 1 Diana Deutsch's "speech-to-song illusion," a perceptual illusion in which, upon repeated listening, small snippets of speech begin to be perceived by the listener as musical.[22] "Upon repeated listening" is key to that definition. It is only through multiple hearings that listeners begin to focus on the melodic contour and the rhythmic content. Deutsch and her colleagues suggest that, upon first hearing, listeners are more focused on the characteristic of the speech stream that are essential for gathering meaning, particularly the identification and

---

[19] Elizabeth Hellmuth Margulis, *On Repeat: How Music Plays the Mind* (New York, NY: Oxford University Press, 2014), 1–3.
[20] Ibid., 5.
[21] This is not to say that repetition is unimportant to language communication. Consider that when students are taught to write, they are encouraged to highlight themes through recurring terms and ideas. But if a student started their introductory and concluding paragraphs with the same sentence—the way the first theme sounds in both the exposition and recapitulation of sonata form—the teacher would likely tell the student that this is unacceptable.
[22] Diana Deutsch, Trevor Henthorn, and Rachael Lapidis, "Illusory Transformation from Speech to Song," *The Journal of the Acoustical Society of America* 129, no. 4 (April 2011): 2245–52.

processing of vowels and consonants.[23] As they re-listen, however, they have the chance to focus on the paralinguistic aspects of the recorded speech, particularly cues like intonation, tone, and articulation.

Steve Reich's early tape works *It's Gonna Rain* (1965) and *Come Out* (1966) are two excellent examples of something akin to the illusory transformation of speech to song. Reich, being a composer of music, casts these manipulated recordings of speech as music and therefore purposefully encourages the listener to pick out the musicality of the speech sounds. On the creation of *It's Gonna Rain*, Reich has said,

> I was extremely impressed with the melodic quality of [speaker Brother Walter's] speech, which seemed to be on the verge of singing. Early in 1965, I began making tape loops of his voice, which made the musical qualities of his speech emerge even more strongly. This is not to say that meaning of his words on the loop, 'it's gonna rain,' were forgotten or obliterated. The incessant repetition intensified their meaning and their melody at one and the same time.[24]

This provides an interesting insight into both the composer's choice of speech as a compositional object and his own listening practice. In this case Reich suggests that his chosen speech segments sound musical to him, and that their repeatability through recording technology is a necessary condition to the structure of the composition—without it, he would not have been able to craft the loops that brought forth the musicality of the speech segment. It is also clear that Reich values the semantic meaning of the speech as much as its phonological content: his segments are based on comprehensible linguistic units. Put another way, the spoken phrases form the basis for musical phrases. Going forward, I want to consider not only the sound of the composition but also the effect that Reich's setting of this "text" has on its meaning and interpretation. I turn now

[23] Ibid., 2251.
[24] Steve Reich, *Writings on Music, 1965–2000* (Oxford University Press, 2002), 19.

to an analysis of the musical structures that emerge from Reich's speech loops, sketching Reich's compositional strategy and how it ties into the sound and meaning of speech.

Because the compositional strategies used in these two works are quite similar, I've chosen to analyze only one of the pieces in-depth: *Come Out.*[25] The work begins with the recorded voice of Daniel Hamm, an African American man who had been the victim of a beating by police during the Harlem riots in 1964. Police were only permitting victims with visible bleeding to be taken to the hospital, so Hamm recounted the following quote, which becomes the basis for Reich's work: «I had to, like, open the bruise up and let some of the bruise blood come out to show them». Reich presents this recorded sentence twice before cutting it down to the shorter phrase «come out to show them» and looping this snippet in two channels. The loops go out of phase over the course of the work, and these new combinations of the repeated segment yield various, subtly shifting polyphonic results.

Although there are two loops, separated in stereo space, that run simultaneously, it is difficult to hear a distinction between the two. What occurs instead is another kind of division, one in which «come out» and «to show them» emerge as distinct entities. I almost always find myself attending primarily to «come out», perhaps because it's the title, perhaps because it comes first in the spoken phrase, perhaps become it could stand on its own semantically as an imperative whereas «to show them» is so clearly in need of a main clause.

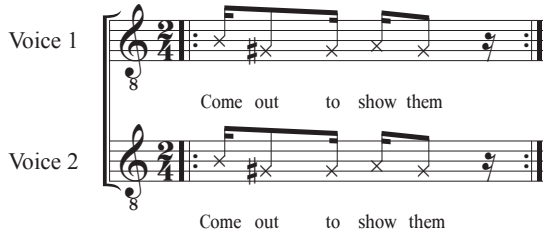
Figure 2.4 shows a transcription of what I hear at several moments in the piece. I don't assert that these are definitively how one should hear this piece. In fact, I think part of what has drawn so many listeners and scholars to this piece is that it allows for so many different

[25] Reich, *Early Works*, CD (New York: Nonesuch, 1987).

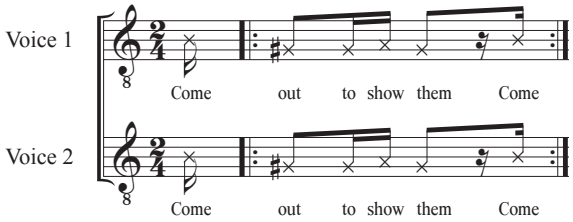
hearings.[26] I claim only that these are illustrative of one possible way through the piece and that by setting down my hearing we can begin to tease out what makes this speech sound musical.

Figure 2.4. Steve Reich, *Come Out*, transcription of my hearing of selected moments

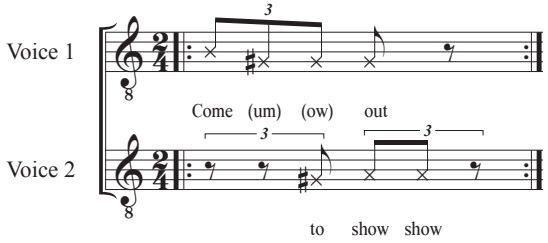
a. 0:13 hearing 1



b. 0:13 hearing 2



c. ~4:00



I'll begin with 0:13. After two times through the complete sentence «I had to, like, open the bruise up and let some of the bruise blood come out to show them» (0:00–0:12), this is where Reich begins to loop the shorter segment «come out to show them». Depending on the listening, I hear two different musical interpretations here, shown in Figure 2.4a and 2.4b. In both the melody and rhythmic durations are the same, but where I hear the downbeat occur changes

[26] For a perspective that casts *Come Out* as a sonic memorial to the Harlem Six case, see Sumanth Gopinath, "The Problem of the Political in Steve Reich's *Come Out*," in *Sound Commitments: Avant-Garde Music and the Sixties*, ed. Robert Adlington (Oxford: Oxford University Press, 2009), 121–144. On the conflict between the Black voice and white composer, see Lloyd Whitesell, "White Noise: Race and Erasure in the Cultural Avant-Garde," *American Music* 19, no. 2 (2001): 168–189. For a psychoanalytical perspective on the role of the listener in Reich's phase works, see David Schwarz, "Listening Subjects: Semiotics, Psychoanalysis, and the Music of John Adams and Steve Reich," *Perspectives of New Music* 31, no. 2 (1993): 43–46. On Reich's tape works as violence against the voice, see Connor, "The Decomposing Voice of Postmodern Music" and Beirens, "Voices, Violence and Meaning."

between the two. The hearing in Figure 2.4a favors an interpretation in which the first syllable is the downbeat. This is, I think, my natural inclination—to hear the first sound as the first beat. It is only as the phrase progresses that my second possible hearing shown in Figure 2.4b emerges. In this interpretation, the longer durations on the works «out» and «them» are afforded emphasis by falling on the beats.

Imbued in both of these moments is the sense of a musical metrical framework. The repetition of the spoken phrase provides a sense of periodicity—because the same short snippet of sound keeps repeating, we feel some sort of pulse associated with each restart of the loop. But a pulse in and of itself is not sufficient to create a sense of meter: Justin London, following Maury Yeston, suggests that a metrical framework must include at least two levels of hierarchy, in this case the beat and the measure.[27] In my hearing, the looped speech forms a measure divided into two beats. The longer rhythmic durations on «out» and «them» are evenly spaced within the perceived measure, which creates a sort of metrical accent, "a beat that is relatively strong in its metrical context," on those words.[28]

This metrical framework remains in place as I listen, and provides a filter through which to hear rhythm when new polyphonic combinations arise as the loops go out of phase. Figure 2.4c demonstrates one such moment. At around four minutes into the piece, the spoken words in the second loop slot evenly between those in the first. My resulting hearing, shown in Figure 2.4c, shows a triplet division of the beat. The «come out» motif is foregrounded. I clearly hear «come» on beat 1 and «out» on beat 2, but the syllables in between blur together to create a sort of

[27] Justin London, *Hearing in Time: Psychological Aspects of Musical Meter* (New York: Oxford University Press, 2004), 17–18. London supports his argument in this section using Maury Yeston, *The Stratification of Musical Rhythm* (New Haven: Yale University Press, 1976), 76.
[28] Fred Lerdahl and Ray S. Jackendoff, *A Generative Theory of Tonal Music* (Cambridge, Mass.: MIT Press, 1983), 17.

stuttering effect («come-um ow-out»). I must attend carefully to hear the «to show them» motif, and when I do it sounds metrically and perceptually weaker than «come out» motif. «Show» falls strongly on beat 2, with «to» forming an anacrusis. Instead of hearing «them», I attune to the «show» from the second loop, perhaps because of the forceful attack of the sibilant /ʃ/ ("sh") phoneme.

This "mishearing" of the linguistic information begs the question, how meaningful are the words being said? Are they repeated to the point where they lose their meaning and become musical sound? *It's Gonna Rain* and *Come Out* are perhaps two of the most recognizable and most analyzed works in the repertory of pieces that use recorded speech as a compositional object, and a survey of the literature on these works reveals the wide range of answers scholars have given to these questions. Maarten Beirens suggests that "the incorporation of pre-recorded speaking voices in [pieces like Reich's tape works] dramatically opens up the musical realm to include social and cultural signifiers that are inextricably connected with the speaker in question."[29] Lloyd Whitesell critiques the racial politics of these works, stating,

> The black voices [in these works] are melodious and expressive, occupying the position of dramatic subjects. The composing persona attempts to disappear within an attitude of rigorous objectivity. But as the musical process is set in motion, the snippets of speech are treated as manipulable voice-objects… Gradually, through phase distortion, semantic articulation and expressivity are utterly drained from the voice-objects. They slowly alter from foreground performers to background noise. It becomes clear that the true 'subject' of the music is the [composer].[30]

Connor, for his part, points out the "disturbing congruity" between the semantic content of the voice and the musical processes used in *Come Out*: "The voice that describes the effect of a beating is itself subjected to a horrifying kind of pulverization; the bleeding described is reproduced in the

[29] Beirens, "Voices, Violence and Meaning," 213.
[30] Whitesell, "White Noise: Race and Erasure in the Cultural Avant-Garde," 177.

effect of bleeding between the different loops of sound."[31] What these scholars seem to suggest is that the meaning of individual spoken phrases themselves don't matter so much as the subjectivity of the speaker behind them. They suggest that (perhaps in conjunction with some knowledge of the origins of the piece and its date of composition) the voice we hear may be that of an African American man, in the 1960s—a time period marked by racial violence and the fight for civil rights in the United States. Even with a degradation of semantic meaning, the expressive features of the voice, the phonological sounds, are tied to meaning.

This encounter with *Come Out* shows us that meaning is made up of more than just the semantic aspect of speech. The sound of the voice itself, emitted from a human body, carries expressive traces of the person who is speaking. The speaker's subject position can play a role in the listener's construction of meaning. Furthermore, Reich's use of repetition via tape loops constructs a musical hearing of the voice, adding a layer of musical meaning that melds with the semantic and expressive meanings of the voice. This piece in particular leaves a lot of room for the listener to make her own judgments about what *Come Out* means and expresses, but such personal interpretations will leverage these three types of meaning.

It is not only direct repetition of short snippets of speech that contributes to the musicalization of speech sounds. Musical repetition at a higher level of the formal hierarchy can also be found in musical works for recorded speech, one involving indirect repetition at the phrase level. To illustrate this, I turn now to another piece by American electro-acoustic composer Pamela Z: *Pop Titles "You"* (composed 1986, recorded 2004).[32] The source text that Z uses for this work is a list of song titles from a record store reference book—all beginning with "You"—a choice that cedes some of her compositional control to chance. This is in some ways

[31] Connor, "The Decomposing Voice of Postmodern Music," 479–80.
[32] Pamela Z, *A Delay Is Better than a Disaster*, CD (Starkland, 2004).

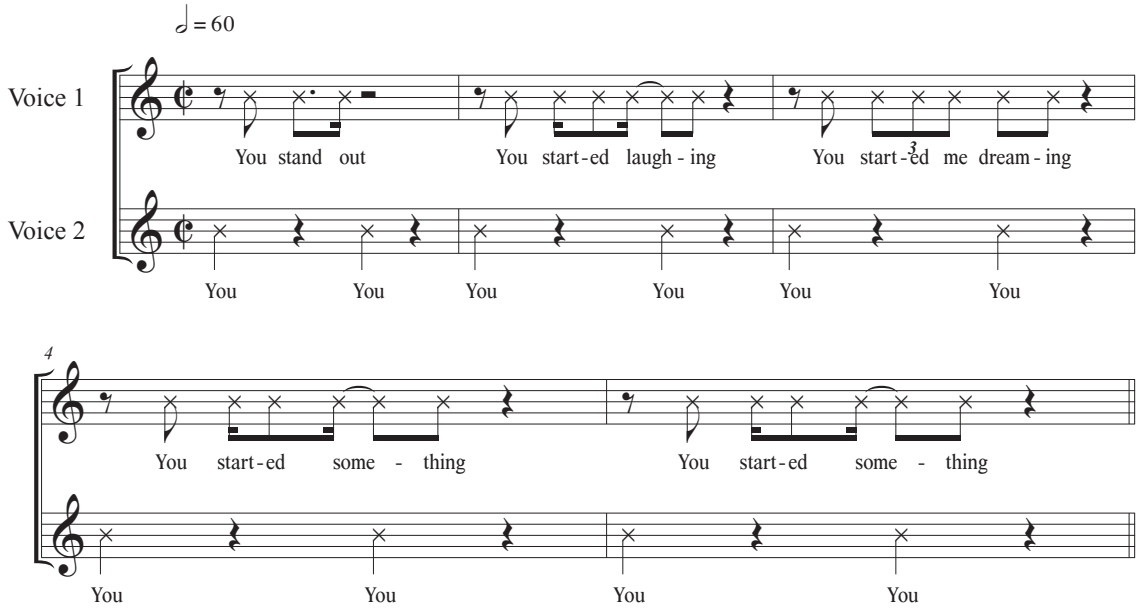
confining—given this approach Z cannot cut phrases that may sound perceptually less musical. In other ways it means that she has more control than Reich over delivery the words—she is the speaker, reading aloud from the catalog of song titles. This ability to regulate delivery is a luxury not afforded to Reich because the genesis of his compositions are found *speech* spoken by another person rather than a found *text* that is later recorded as speech.

Another notable divergence from Reich's *It's Gonna Rain* and *Come Out* is that the foregrounded recorded speech in *Pop Titles "You"* is not a tape loop of a single short spoken phrase but a continuous delivery of individual song titles. The repetition here is less direct than Reich's but nonetheless important to the musical structure: because each song title begins with the word "you," there is a feeling of repetition, something that marks the beginnings of phrases and sets up a sort of motive by which we can tie together spoken phrases in a musical way.

Before considering the effect that repetition has at the phrase level, let me first explore how Z sets up the musical metrical framework—a temporal framework in which time is divided into regularly recurring patterns—by means of exact repetition. A loop of the word "you" sounds every second, setting a tempo of 60 bpm, a metric scaffold upon which to interpret the spoken phrases. In the introduction (0:00–0:12), the regularly-repeated «you»s form a minimal temporal framework; they do not form a hierarchically-organized meter. According to Justin London, "periodic stimuli," such as these regularly occurring «you»s, "can (and probably do) trigger a metric response in the listener, setting up a dynamic framework against which subsequent durations or onsets may be heard and measured."[33] The regularly-repeated «you»s form a tactus.

[33] Justin London, *Hearing in Time,* 13.

Figure 2.5. Pamela Z, *Pop Titles "You,"* transcription of first five spoken phrases (0:12–0:22)

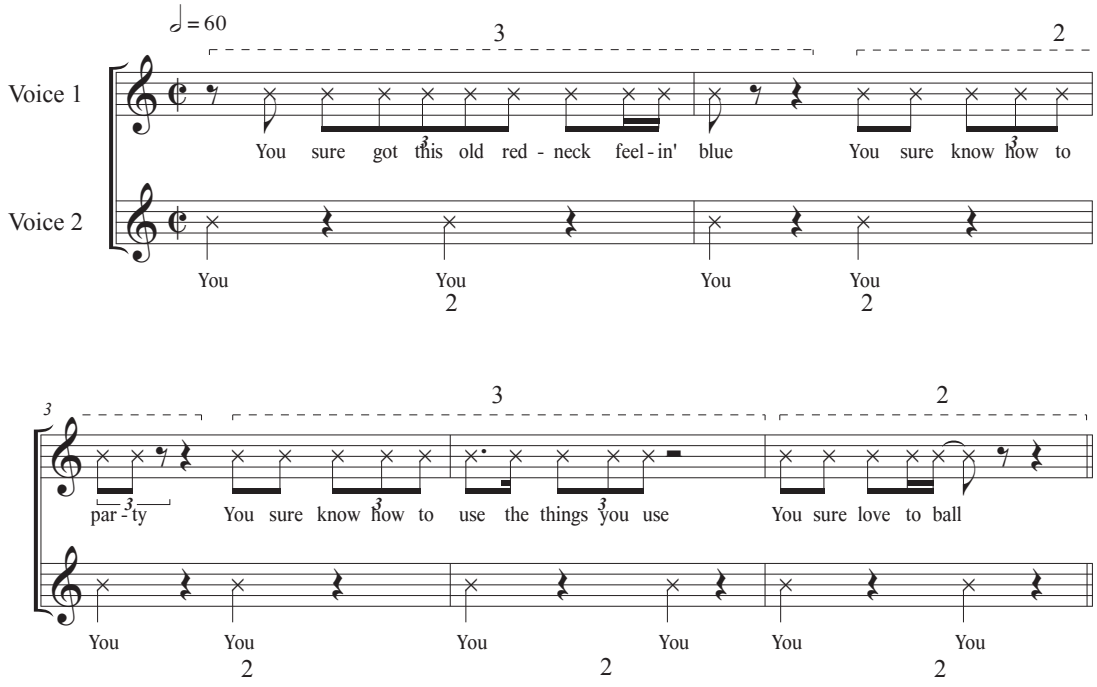


It is only when the spoken phrases begin (0:12), however, that a sense of metrical hierarchy starts to emerge. Figure 2.5 shows a transcription of the first five spoken phrases. Each phrase occurs over the space of two of the beat-layer «you»s, creating a feeling of cut time by setting up an expectation that these individual events, grouped together by semantic content—a song title—and delimited by silence on either side, will occur at regular metric intervals.

As can be seen in Figure 2.5, each spoken phrase begins immediately after the downbeat. Shortly after, the metrical position of the beginnings of phrases changes so that they fall on the beat (0:24–1:17). This serves to strengthen the established feeling of duple meter. At times, however, spoken phrases chafe against the confines of the metrical framework. At 0:46, the longest spoken phrase so far («You sure got this old redneck feeling blue») takes place over the span of three beats instead of two, introducing a moment of what Harald Krebs has called metric dissonance. Figure 2.6 shows this metric dissonance through the relationship between metrical layers. The metrical layer of «you»s remains unchanged while the antimetrical layer of spoken phrases alternates between groupings of three and two beats. The three-beat grouping

dissonance—a type of metrical dissonance that is formed when two interpretive layers of metrical material have different cardinalities and are not multiples/factors of each other (in this case groupings of three beats vs. two beats).[34] This first dissonance sets up a second kind of dissonance in the second phrase: a displacement dissonance, which results from the different positioning of congruent layers of metrical material.[35] A second three-beat grouping dissonance is then necessary to correct this displacement dissonance. Finally, the fourth phrase of this example returns to metric consonance. Here, metric dissonance and its resolution integrate spoken phrases of varied lengths into the metrical framework.

Figure 2.6. Pamela Z, *Pop Titles "You,"* metric dissonance (0:46–0:56)



[34] Harald Krebs, *Fantasy Pieces: Metrical Dissonance in the Music of Robert Schumann* (New York: Oxford University Press, 1999), 31.
[35] Ibid., 33.

As purely a spoken vocal work, Z's piece is fairly simple: there is no musical pitch system that the speech must confront, the beat layer sounds only on the pulse, spoken phrases for the most part fall neatly into the space of one measure, and whatever metric dissonances that do arise are quickly resolved. Nonetheless, *Pop Titles "You"* provides us some insight into how speech and musical features may interact with one another and how repetition plays into this interaction. The most basic musical building block in this equation is the pulse-level «you», here explicitly repeated at 60 bpm. This tactus forms a basic temporal framework against which all speech rhythms can be judged musically. The importance of the relative uniformity of phrases in Z's source material is notable: phrases are often of a similar syllabic length—usually five syllables as in Figure 2.5—which results in a similar temporal duration. The repetitive linguistic structure of the spoken phrases lends itself to the formation of musical phrases, which in turn contributes to large-scale rhythmic expectations. Each spoken phrase begins with the subject «you» (which is roughly the same duration each time) followed by at least two—but often four—additional syllables. Because the spoken phrases form a semantically understandable unit and are offset on either side by silence, they are easily parsed as formal units in the metrical framework. For the most part, due to similar syllabic lengths the phrases recur at regular intervals, setting up the measure in the metrical framework. At times, however, the spoken phrase is temporally longer and semantically more complex, requiring that it exceed the bounds of the musical measure. In such cases metric dissonance arises on the musical surface.

The similarity of linguistic phrase structure, in addition to the aforementioned effect of its phonological structure, has an effect on the way the listener processes the semantic content. Hazel Smith, in her chapter on the voice in computer music, notes that *Pop Titles "You"* "highlights one of the most seductive devices in poetry (the use of the second person and the

ambiguity it creates about the addressee)."[36] The repetition—both of the direct address "you" and of the speech phrase as musical unit—has a sort of hypnotic effect. While it is clear that in terms of semantic content the spoken phrases are only tangentially related, the pronoun "you" lends a kind of immediacy to the utterance, drawing the listener's attention by seeming to address her directly. This immediacy is highlighted by the production and vocal delivery on the recording: Z's voice comes across as soft and somewhat breathy. Her voice has been close miked, isolating the signal from any ambient room noise or reverberation (some of which is added in in post-production).[37]

Because of the repetitive nature of the phrase structure, I find myself as a listener often tuning out the voice's meaning throughout the litany of «you» phrases, but certain phrases that seem out-of-character for Z's voice jar my attention back to the semantic meaning of the linguistic utterances. For example, when I hear Z's voice say «you sure got this old redneck feeling blue», I'm drawn to the incongruity of Z's decidedly non-redneck-like pronunciation. When she says «you sure tell it like it is, George Jones», it strips away the artifice of the pronoun-as-direct-address. Previous phrases create a sense that only two persons exist in this monologue: the speaker and the "you" she addresses. By naming the subject (George Jones) this phrase introduces a third person, which takes away some of the intimacy of address to "you," the listener. While the recorded voice in this piece does contain understandable, semantically and syntactically correct phrases, to me the content of the voice serves more to structure the musical

[36] Hazel Smith, *The Contemporary Literature-Music Relationship*, Routledge Interdisciplinary Perspectives on Literature 65 (New York: Routledge, 2016), 27.
[37] For more on the use of the microphone as an expressive device in music, see Martin Stokes, "'Abd al-Halim's Microphone," in *Music and the Play of Power in the Middle East, North Africa and Central Asia*, ed. Laudan Nooshin (Farnham: Routledge, 2009), 49–61; Idem., *The Republic of Love: Cultural Intimacy in Turkish Popular Music* (Chicago: University of Chicago Press, 2010), 61–62; and Nicola Dibben, *Björk* (Bloomington: Indiana University Press, 2009), 107–09.

surface than to express linguistic meaning. One might even argue that similarity between the song titles serves to highlight the redundancy or vapidity of the lyrical themes in pop music. Ultimately, repetition in *Pop Titles "You"* brings to the fore similarities between spoken phrases that point to both musical and linguistic motifs.

Both Reich's *Come Out* and Z's *Pop Titles "You"* demonstrate the role that repetition has in the musicalization of speech. As is quite clear in these examples, repetition, as a temporal event, provides a basis for a metrical framework. Both works rely on short loops («come out to show them», «you») to create a regularly recurring event that forms the basis for periodicity. The rhythms of speech are then perceptually musicalized according to this framework. While Z's "repetition" is less exact than Reich's (relying as it does on newly uttered phrases), similarities in phrase structure make it possible to construct a metric framework against which the listener can hear deviations from this established pattern as dissonant. By manipulating the spoken voice by way of technologically-mediated repetition, these composers can use the voice to establish temporal expectations that are systematically organized in a way that sounds musical. Moreover, such repetition encourages the listener to attend to the phonological content of speech—particularly its pitch and durational qualities—over its semantic meaning, shifting the purpose of the speech from communication to sound.

2.3. Technological manipulation

We've seen some level of technological manipulation in each of the previous examples in this chapter—all of the aforementioned Reich pieces use tape to splice and loop portions of the original speech recoding, to segment and repeat, and Z employs looping, digital delay, and live processing to craft *Pop Titles "You."* The usage of recording technology is an essential element of the music in this corpus as it opens up the possibility to replay an exact "performance" of the

source speech. So far we've attended to only snippets of recorded speech that are relatively faithful reproductions of the speaker's delivery. But beyond using technology for direct repetition and for segmentation, some composers employ additional technologies in order to change the pitch, rhythmic, and/or timbral qualities of the recorded speech.

I turn first to the Gregory Brothers, a musical group known best for their YouTube video series *Auto-Tune the News* and *Songify This*, in which speech recordings are musicalized, usually in order to highlight idiosyncratic manners of speaking. The Gregory Brothers digitally manipulate recordings of speech via pitch-shifting software and set these to musical accompaniment in various popular genres. If you're familiar with viral videos, you've likely seen their "BED INTRUDER SONG!!!!,"[38] which has garnered more than 147 million views since it was uploaded in 2010. In this video the speech from the source recording, a newscast interview with Antoine Dodson, an eyewitness to a home invasion, is auto-tuned over a hip hop beat. The success of "Bed Intruder" led the Gregory Brothers to create more videos that poke fun at the speech found in a broad swath of American popular culture, including samples from politicians, memes, and popular television shows.

This format was later parodied in the theme song for Netflix sitcom *Unbreakable Kimmy Schmidt* (2015–2019), which was created by film/television composer Jeff Richmond in collaboration with the Gregory Brothers.[39] This song follows the same basic structure as the Gregory Brothers' first viral hit, "Bed Intruder." The "source material" is a scene in the show in which a poor Black man who witnessed a crime in his neighborhood is interviewed for a news

[38] schmoyoho, "BED INTRUDER SONG!!!," YouTube, July 31, 2010, https://www.youtube.com/watch?v=hMtZfW2z9dw.
[39] While the show features a shortened version of the song, the full recording can be found at schmoyoho, "Unbreakable Kimmy Schmidt - Songify This! (Theme Written by Jeff Richmond)," YouTube, March 6, 2015, https://www.youtube.com/watch?v=WYNbp0u8WjA.

broadcast. The interview subject, played by actor Mike Britt, has animated gestures and a distinctively expressive voice. As journalist Kevin McFarland writes, "In retrospect it's hard to listen to Britt's inflections without hearing the musicality in it, and [show creators Tina] Fey and [Robert] Carlock's edit heightens that impression."[40] Ultimately the compositional process was to pull out features of Britt's delivery that seemed to go beyond everyday speech and to set them to music that suited the mood of the show. Jeff Richmond described the composition of the theme song in an interview with the *Los Angeles Times*:

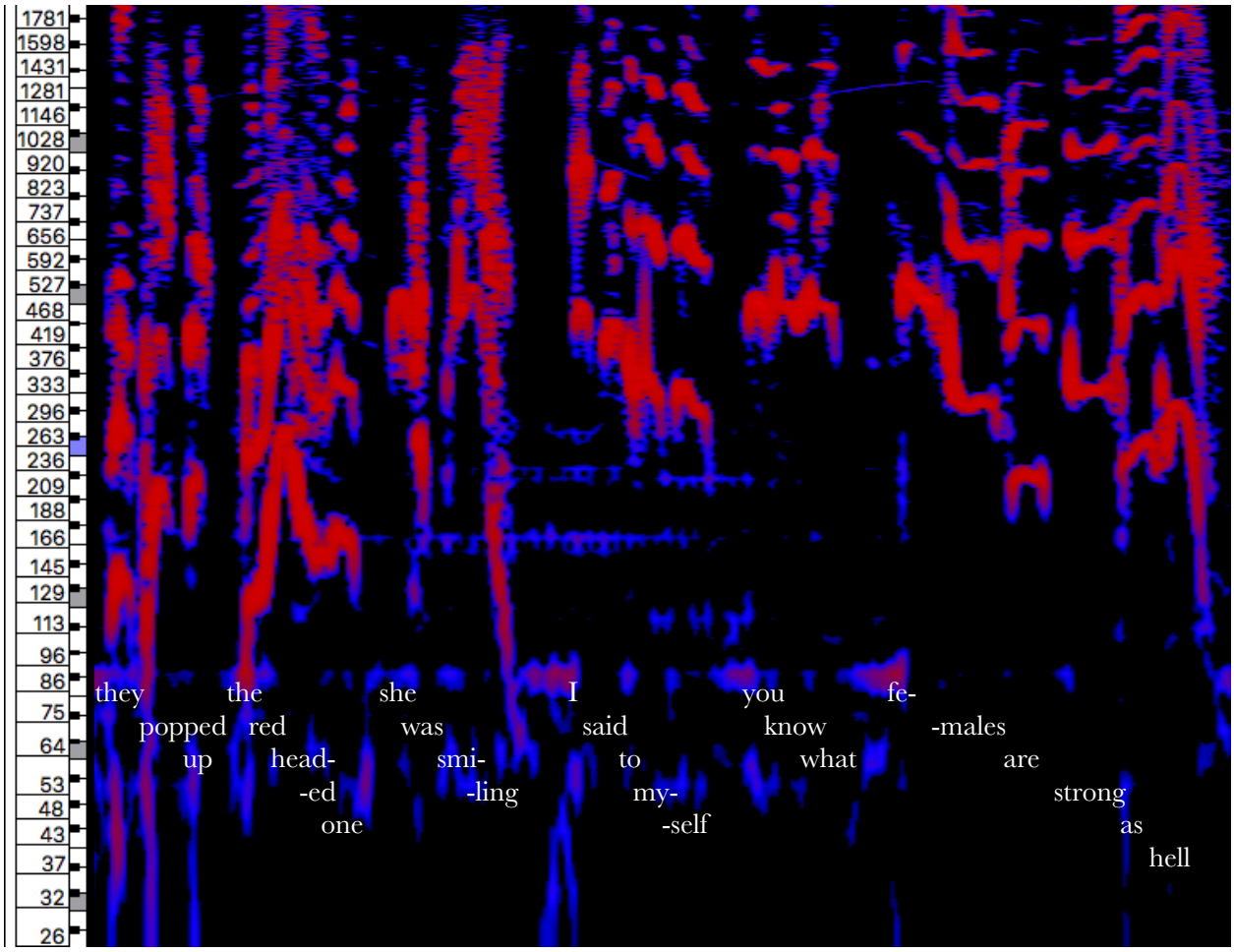
> I think they [Fey and Carlock] wanted it [the theme song] to be [upbeat]…because we were going to witness a really terrible, grim backstory for our lead character. We needed something to be like, "OK, but it's still going to be fun and bright and sunny, and funny."…Tina and Robert kind of crafted what the monologue was going to be, and we just pulled out these expressive, declarative statements to train the song around. …We took that and we did our own Auto-Tune and built the chord structures around it. And then we added some female singers to it to make it more femme-power and to heighten the nature of the whole thing.[41]

This suggests that the ultimate goal of the song is to communicate to the listener—both musically and linguistically—the basic background for the show and its general tone: for years the optimistic and energetic protagonist Kimmy Schmidt was held captive by a cult leader in an underground bunker in Indiana and upon her rescue she's determined to move past her horrific experience by moving to New York City. The following analysis will delve into *how* the song communicates these ideas through its musicalization of speech.

---

[40] Kevin McFarland, "The Secret Behind Kimmy Schmidt's Perfect Theme Song," *Wired*, March 12, 2015, https://www.wired.com/2015/03/kimmy-schmidt-theme-song/.
[41] Yvonne Villarreal, "'Kimmy Schmidt' Composer Jeff Richmond on the Show's Viral-Friendly Songs," latimes.com, March 31, 2015, https://www.latimes.com/entertainment/music/posts/la-et-ms-kimmy-schmidt-jeff-richmond-peeno-noir-20150327-story.html.

Figure 2.7. Spectrogram of «they popped up, the redheaded one, she was smiling! I said to myself, 'you know what? Females are strong as hell!'», "Unbreakable Kimmy Schmidt | Bankston Interview | Netflix," 1:18–1:27



As noted, the source material is actor Mike Britt's recorded speaking voice, a voice chosen for its lyrical qualities.[42] Andrew Gregory, guitarist and vocalist for the Gregory Brothers, praised Britt's delivery, saying that his voice has "both an amazing sonic quality somewhere in between

[42] While we have seen in this chapter mostly source recordings chosen for their inherent "musicality" of speech (Reich's *Come Out* and *It's Gonna Rain*, later Lansky's *Six Fantasies on a Poem by Thomas Campion*), this is not always the case. For example, as I will argue in the next chapter, many of the recorded speaking voices in Peter Ablinger's *Voices and Piano* are not easily musicalized.

Joe Cocker and James Brown, and also the really powerful emotional delivery."[43] We can hear this in, for example, the portion of the original monologue that climaxes with «females are strong as hell» (Figure 2.7).[44] Even a cursory glance shows the melodic contour rising dramatically over the course of the excerpt only to fall in the last phrase. But beyond that, we can also hear a dramatic rise-fall contour over the word «smiling», which emphasizes the exclamatory nature of this final word of a sentence. The delivery of the final phrase, «females are strong as hell» (which becomes a particularly important motive in the theme song) contrasts with the phrases leading up to it in a number of ways. The preceding phrases («they popped up», «the redheaded one», «she was smiling», «I said to myself», «you know what») are delineated by clear breaks at the end of each phrase while the phrases themselves are smoothly connected. This final phrase is marked not only by its falling pitch contour and by being the longest phrase so far, but by Britt's sort of staccato articulation of individual syllables.

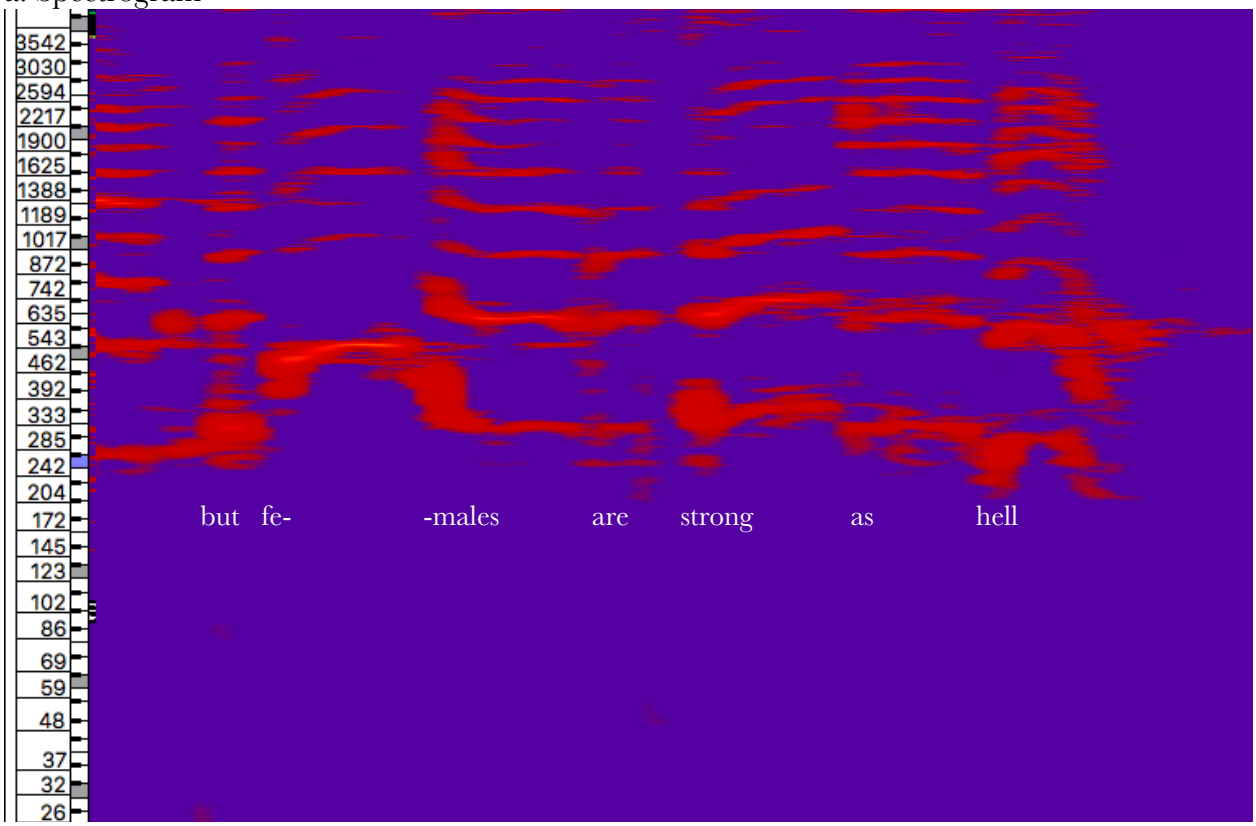
In the context of the Kimmy Schmidt theme song, Britt's monologue is spliced and reordered, its pitches and rhythms slotted into a tonal, metrically-regular musical framework by means of digital alteration. The last line shown in Figure 2.7, «females are strong as hell», provides a good example of the effect of these alterations. Not only is this phrase a particularly climactic point in Britt's monologue, but it becomes something of a refrain in its musical context, occurring in the lyrics both at the end of the first verse and twice in the chorus, in a similarly important position as the final line in a stanza. The lyrics illustrate something of a recurring theme in the show—Kimmy's resilience in overcoming her trauma.

---

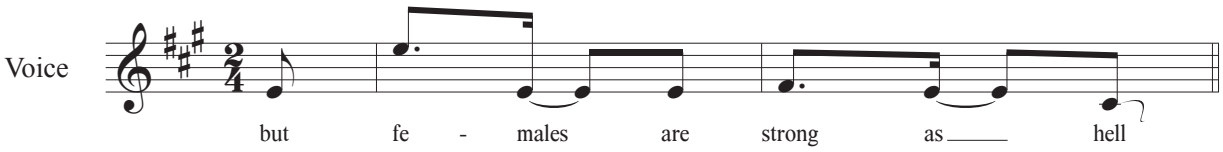
[43] McFarland, "The Secret Behind Kimmy Schmidt's Perfect Theme Song."
[44] Netflix, "Unbreakable Kimmy Schmidt | Bankston Interview | Netflix," YouTube, March 9, 2015, https://www.youtube.com/watch?time_continue=30&v=HUk6GmqfqnY, 1:18–1:27.

Figure 2.8. «but females are strong as hell», Jeff Richmond and the Gregory Brothers, "Unbreakable Kimmy Schmidt - Songify This! (Theme Written by Jeff Richmond)," 0:39–0:42[45]

a. Spectrogram



b. Transcription



As we can see in Figure 2.8, while the line «females are strong as hell» is somewhat smoothed out through the resources provided by AutoTune it retains some of the natural features of Britt's delivery. The «fe-» syllable is still higher than «-males», giving particular emphasis to the first syllable of the word (and the phrase). There is still a descending melodic contour, albeit

[45] To create this figure the vocals have been isolated from the musical accompaniment using PhonicMind.com, an online vocal remover.

one in which individual syllables maintain a steadier fundamental frequency. The exception is the word «hell», which still maintains its rising-then-falling contour, retaining a markedly expressive moment in Britt's delivery.

This example demonstrates one way in which composers might use digital manipulation to strike a compromise between musical frameworks and the original expression of the speaker, and contrasts with the musicalization of speech in Reich's *Different Trains*. In *Different Trains*, the pitch and temporal qualities of the recorded speech are unaltered and the musicalization is accomplished by the *suggestion* of musical rhythms and pitches in the strings. Here Richmond and the Gregory Brothers' musicalization of speech involves actually altering the recording itself so that the sound of the speech itself fits the musical pitch and rhythmic framework established by the accompaniment. Richmond and the Gregory Brothers use the unique melodic contour of Britt's delivery as a starting point, and then employ digital means to "smooth out" the pitches so that they fit into this duple meter, tonal soundworld.

I turn now to American composer Paul Lansky's (b. 1944) *Six Fantasies on a Poem by Thomas Campion* (1978–79) in order to demonstrate the musicalization of speech through technological manipulation with more abstract results—Lansky takes the soundworld of the speech recording much further from their original form than Richmond and the Gregory Brothers. In *Six Fantasies*, Lansky's source is a recording of the actress Hannah Mackay, Lansky's wife, reciting a poem from English composer and poet Thomas Campion's (1567–1620) treatise *Observations in the Art of English Poesie* (1602). Lansky himself describes Campion's treatise as an attempt

> to construct a quantitative model for English poetry in which, as in ancient Latin and Greek, meter is determined by vowel quantity, rather than by accentual rhythm. The poem [on which *Six Fantasies* is based] is a wonderful, free-wheeling spin about the vowel box. It is almost as if he is playing vowels the way one would play a musical instrument, jumping here and there, dancing around with dazzling invention and brilliance, carefully balancing repetition and variation. The poem itself is about Petrarch's beloved Laura, whose beauty expresses an implicit and heavenly music, in contrast to the imperfect, all

too explicit earthly music we must resign ourselves to make. This seemed to be an appropriate metaphor for the piece.[46]

Campion includes this poem as an example of lyrical poetry that may easily be set to music, "fit for Ditties or Odes":[47]

Rose-cheekt Lawra, come
Sing thou smoothly with thy beawties
Silent musick, either other
Sweetly gracing.

Lovely forms do flowe
From concent devinely framed;
Heav'n is musick, and thy beawties
Birth is heavenly.

These dull notes we sing
Discords neede for helps to grace them;
Only beawty purely loving
Knows no discord;

But still moves delight,
Like cleare springs renu'd by flowing,
Ever perfect, ever in them-
selves eternall.[48]

Lansky uses a recording of Mackay's recitation of the poem as the basis for his composition, manipulating her speech by way of all-pole linear prediction, a music synthesis technique developed by Lansky over the course of the composition of this work.[49] This technique allowed Lansky to analyze the speech recording in 1/112th-of-a-second segments and then reconstruct the

[46] Paul Lansky, *Fantasies and Tableaux*, CD liner notes, CRI 683 (New York: Composers Recordings Inc., 1994).
[47] Thomas Campion, "Observations in the Art of English Poesie [1602]," in *Campion's Works*, ed. Sylvanus Percival Vivian (Oxford: Clarendon Press, 1909), 49.
[48] Ibid., 50–51.
[49] Denise Michelle Ondishko, "Six Fantasies on a Poem by Thomas Campion: Synthesis and Evolution of Paul Lansky's Music Compositions" (Ph.D., University of Rochester, Eastman School of Music, 1990), 16.

original compositional object with independent control over pitch, rhythm, and timbre.[50] The result is that "when resynthesized, the voice gain[s] all the virtuosity and versatility of the computer medium while retaining its human character and presence."[51] *Six Fantasies* takes the human voice far beyond its natural sonic capabilities, but yet Mackay's voice still maintains a level of subjectivity and individuality: the inflections and expressiveness are still distinctly her own. Each of the six movements is concerned with a different sonic aspect of the speaking voice. According to Lansky, the first movement, "her voice," is concerned with pitch contours; the second, "her presence," with the vowels; the third, "her reflection," with "the acoustical resonance of speech"; the fourth, "her song," with "the mouth qualities of speech"; the fifth, "her ritual," with "the percussive aspects of the speech."[52] The final movement, "her self" presents the speech recording in a largely unaltered form (with electronic accompaniment), eventually revealing what Mackay's recorded voice sounds like with minimal technological interference.

In his analysis of *Six Fantasies* David Code argues that the vowel timbres in this work are a structural element that creates unity between the six movements, and focuses his analysis on the frequencies produced by these vowels.[53] While these fundamental and formant frequencies certainly do play an important part in the pitch structure of the work, I hear a number of additional factors (the articulation of consonants, the speed of delivery, and the quality of the voice) as equally important to the structure of the work. In order to explore the effect of these

---

[50] David Loberg Code, "Observations in the Art of Speech: Paul Lansky's Six Fantasies," *Perspectives of New Music* 28, no. 1 (1990): 147.
[51] Elliott Schwartz and Daniel Godfrey, *Music since 1945: Issues, Materials, and Literature* (New York: Schirmer Books, 1993), 385.
[52] Ondishko, "Six Fantasies on a Poem by Thomas Campion," 29. This quote is from a personal interview with Lansky done by Ondishko, and the editorial brackets are hers.
[53] Code, "Observations in the Art of Speech," 145.

phonetic elements on the musical work, I turn now to movement 3, which Denise Ondishko characterizes as an abstract, non-pitched, and highly repetitive movement.[54]

In the third movement, "her reflection," the use of reverberation results in a sort of sonic "washing out" of the semantic content as aural emphasis is placed on the pitches of vowel formant frequencies because they are prolonged. This effect is reminiscent of Alvin Lucier's *I Am Sitting in a Room* (1969), but unlike the eventual outcome of Lucier's piece, in which the resonances of the room completely smooth out the voice into pure tones, there are still audible traces of humanity left in Mackay's voice in *Six Fantasies*. I believe part of this is due to the quality of Mackay's voice and delivery, both of which constitute suprasegmentals—that is, "aspects of speech which persist over several segments, such as duration, loudness, tempo (speed), pitch characteristics and voice quality; they are often thought of as the 'musical aspects of speech.'"[55] As an actress, Mackay has had vocal training that has shaped her manner of speaking. Ondisko describes Mackay's voice as "immediately noticeable: her vowels are very clear and stable and her consonants are distinct. The final result is a speaking voice which is melodious, gentle, and is a pleasure to listen to all by itself."[56] Beyond the unique suprasegmental factors of Mackay's voice, perhaps the most noticeable factor for my hearing of the voice as such in "her reflection" is the articulation of consonants, which are largely still audible and at times even enhanced by the reverberation. The following analysis explores the way in which Lansky retains acoustic information that is unique to Mackay's vocal delivery despite the heavy technological intervention that makes the speech itself difficult to understand.

---

[54] Ondishko, "Six Fantasies on a Poem by Thomas Campion," 29. Contrastingly, movements 1, 2, and 6, which are "sung"—meaning that the pitch content is emphasized by Lansky's processing—with accompaniment. Movement 5 is "sung" without accompaniment.
[55] Richard Ogden, *An Introduction to English Phonetics*, Second Edition, Edinburgh Textbooks on the English Language (Edinburgh: Edinburgh University Press, 2017), 23.
[56] Ondishko, "Six Fantasies on a Poem by Thomas Campion," 23.

Figure 2.9. Spectrogram of «Rose cheekt Lawra», Paul Lansky, *Six Fantasies on a Poem by Thomas Campion*, III. "her reflection," 0:10–0:15

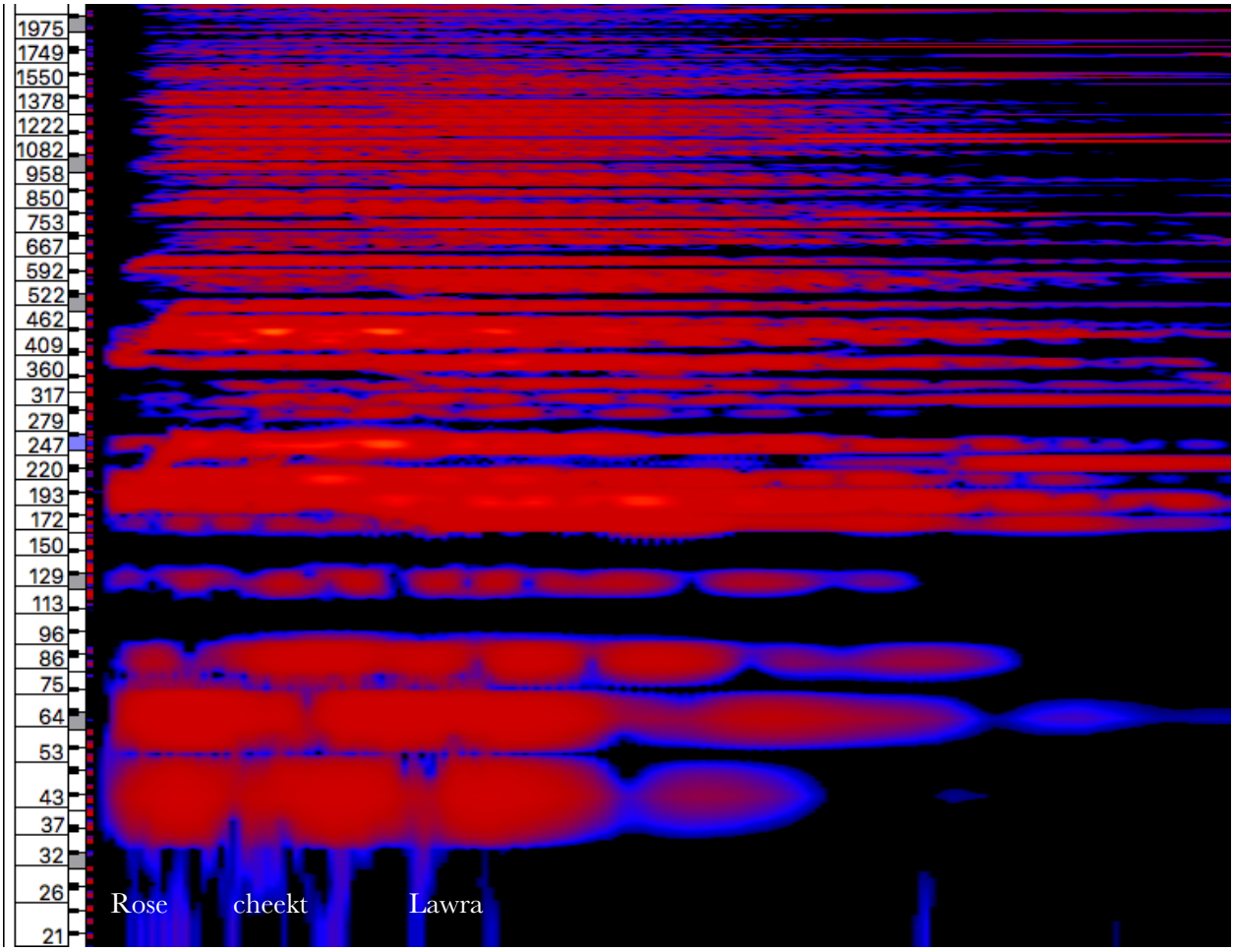
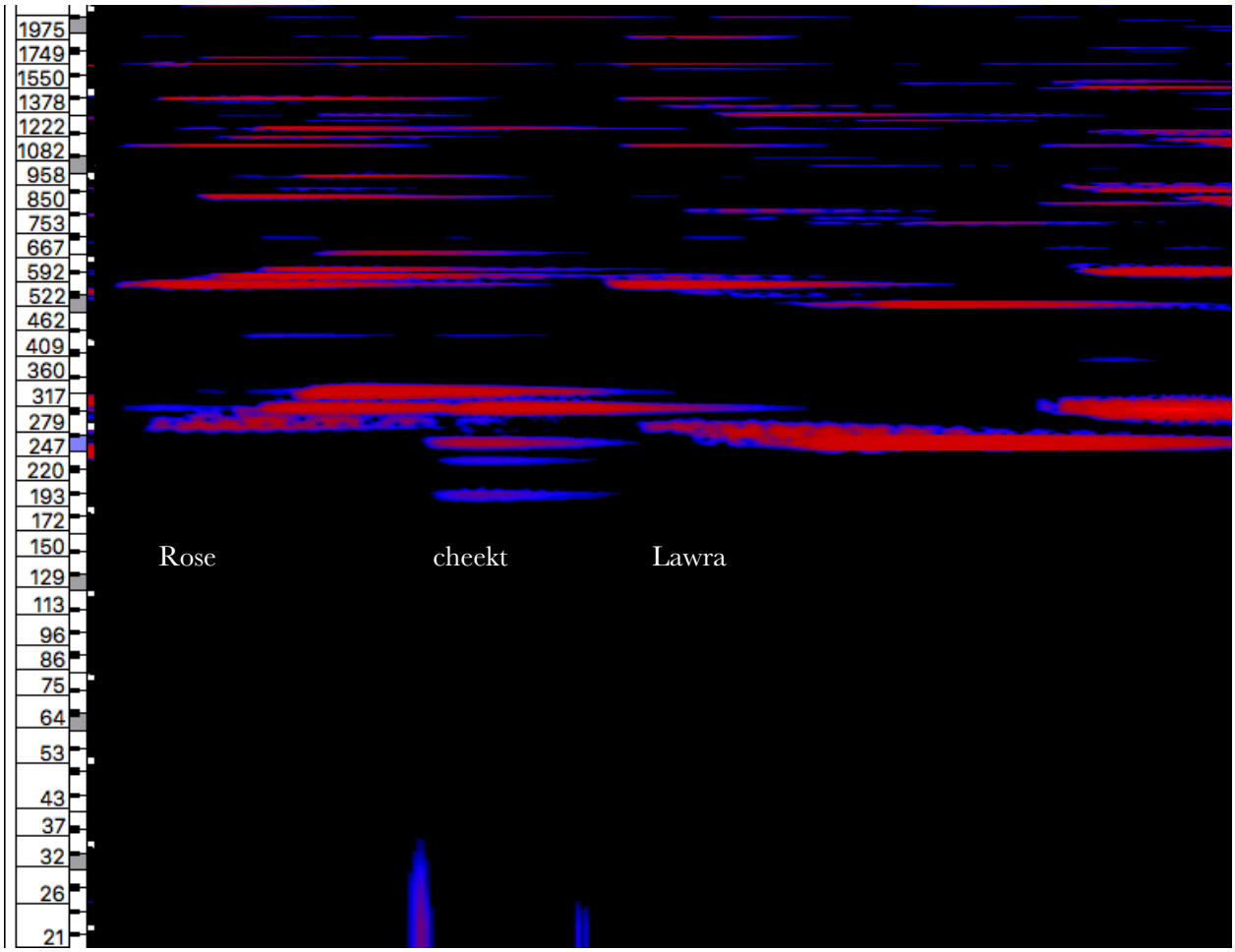


Figure 2.9 shows a spectrogram of the first utterance of «Rose cheekt Lawra» in this movement. Due to the resonances of the reverberation, the frequency content of this excerpt is both quite static and quite dense, with many overtones ringing out. This creates a sort of blending that makes individual articulations of syllables hard to discern (both aurally and visually)—I personally can't really hear the articulation of the second syllable in «Lawra». The sharp articulation of /tʃ/ at the beginning of «cheekt», however, stands out; particularly in the lowest pitch, we can see a sort of disturbance in the pitch: there is a momentary closing off of the sound between «Rose» and «cheekt», and the articulation of the second word produces a fair bit

of noise that we can see below the pitch on the spectrogram. This aural emphasis is due in part to how to this consonant is produced: /tʃ/ is an affricate, which involves a stop closure of the vocal tract followed by a slow release that results in audible friction.[57] In short, it is a noisy consonant, and this noise, when emphasized through reverberation, disrupts the focus on musical pitch.

Figure 2.10. Spectrogram of «Rose cheekt Lawra», Paul Lansky, *Six Fantasies on a Poem by Thomas Campion*, III. "her reflection," 0:23–0:29



[57] Philip Carr, *A Glossary of Phonology*, Glossaries in Linguistics (Edinburgh: Edinburgh University Press, 2008), 10.

Lansky plays with this phrase, setting it quite differently when it occurs again, as shown in Figure 2.10. Again, Mackay's distinct manner of speaking shines through despite its technological transformation. Over the course of the word «Rose», we can hear a rising melodic contour, which Lansky has striated here into three distinct pitches. As with the first utterance of this phrase, the articulation of /tʃ/ at the onset of «cheekt» in noticeable through its lower pitches, although here those have been smoothed out over a longer duration by the reverberation. «Lawra» is vocalized with a melodic descent, making audible the rising-then-falling contour for the suprasegmental phrase that was somewhat buried in the thicker texture of the first utterance. By reducing the number of overtones heard in this second utterance, Lansky allows a bit more of Mackay's voice to shine through, although it is still highly processed.

The focus throughout *Six Fantasies*—both in Lansky's musical work and in the text itself—is on exploring the musicality of the speaking voice. The text of the poem invokes the heavenly aspects of music, and Lansky's use of technology in his composition fittingly makes the speaking voice sound otherworldly. In "her reflection," this otherworldliness is achieved by processing the recorded voice until it is almost unrecognizable. Even so, Mackay's presence continues to shine through from retained features of her vocal delivery. The first utterance is a very restrained, mechanical version of the phrase, sounding almost robotic. The second is sparse, perhaps harder to understand due to its limited pitch content, but through its melodic contour it better captures the humanity of the speaking voice.

As a whole, *Six Fantasies* is an example of how composers can use technology to deeply alter the voice in the quest to bring out its musical features, and yet not completely destroy the individuality and subjectivity of the speaker. I focused here on the third movement because it presents the furthest degradation of the voice in this work—the intelligibility of speech is all but

obscured, yet identifiable features of Mackay's vocal delivery remain. The other movements create a similar effect. In the fifth movement for instance, which Ondishko groups with the third movement as "abstract, non-pitch[ed], [and] highly repetitive," Lansky turns Mackay's voice into a sinister whisper through use of comb filters in the service of emphasizing her consonants.[58] Mackay's speech remains semantically intelligible even as its pitch contours are stripped away. In the first movement, the voice is synthesized at three different pitch levels moving in parallel motion. The multiplication of the voice into three gives it an eerie, otherworldly feel, but the pitch contours and rhythms of speech remain. In all of these movements, Mackay's individual expression shines through despite major alterations to the pitch, timbral, and temporal features of her voice. *Six Fantasies* underscores how unique a human voice is and shows how the speaker exerts force on the musical work even as the composer manipulates the voice.

Conclusion

Ultimately, all of the musical works studied in this chapter involve some level of curation on the part of the composers in the choice of spoken text. In the examples by Reich, and Richmond and the Gregory Brothers, small samples of a longer recorded monologue are chosen for their inherent "musicality." In the examples by Z and Lansky, the text used has its basis in poetry, providing a basis for a heightened state of expression in the speakers' vocal delivery.[59] This curation is enacted by way of some form of technological manipulation—from relatively low-tech looping to state-of-the-art computer processing—that is used to craft the musical structure. Technology here is a means of enhancing the musicality of speech, fitting it into

[58] Ondishko, "Six Fantasies on a Poem by Thomas Campion," 29.
[59] In Z's case the song titles, drawn mainly from the lyrics of the songs themselves, are abstracted from their poetic context. Nonetheless, even divorced from their accompanying rhymes, many of these phrases carry a poetic feeling to them and would be out of place in everyday conversational speech

metrical frameworks or musical pitch systems while still retaining the qualities of the vocal delivery that made the speech recording sound "musical" in the first place.

The alteration of speech into music in contemporary Western music

Theorizing the major issues at the heart of post-war music composition, Elliott Schwartz and Daniel Godfrey argued in 1993 that, "During the past half century composers have tried to extend the expressive resources of the individual performer through two basic approaches: (1) by broadening the range of activities—movement, gesture, vocal utterance—required as part of the performance, and (2) by augmenting the range of sounds that may be produced by an instrument or voice."[60] Indeed, all of the composers considered in this chapter—covering a range from minimalism, electroacoustic, and computer music to television music and meme culture—seem interested in this approach to music. Reich and Richmond specifically remarked on "hearing music" in their subjects' speech. Lansky says that, "rather than trying to liberate our musical perceptions from traditional notions of music, I'm interested in harnessing the world-building power of familiar musical conceptions to enhance our perceptions of the sounds of the world. Often, I take the sounds of the world and impose 'music' on them, or use them as excitation functions for music."[61] These composers are broadening the range of musical expression to include the speaking voice.

The composers are interested in speech, yes, but the role of technology in the production of these musical works is undeniable. I think Z gets close to the heart of the issue when she notes that "some of the most exciting work I've seen [in music composition] lately combines very different types of tools—acoustic instruments with electronic ones, mechanical devices with

[60] Schwartz and Godfrey, *Music since 1945*, 138.
[61] Jeffrey Perry, "The Inner Voices of Simple Things: A Conversation with Paul Lansky," *Perspectives of New Music* 34, no. 2 (1996): 43.

digital devices, machines with flesh and blood instruments."[62] The alteration of speech in the works explored in this chapter all seek to meld the human body with musical tools. Repetition and segmentation are musical devices that have the ability to change the way the recording of speech communications, pushing it into a more musical mode of delivery. Technological manipulation accomplishes a similar goal, having the potential to tease out the pitch and rhythmic qualities of recorded speech and augment them until they sound more traditionally musical, adhering to a 12-note division of the octave and falling neatly into a given meter. These pieces combine the sounds of the human body—the lowest-tech means of producing music (and indeed involving sounds that aren't traditionally considered musical)—with 20th- and 21st-century technologies for sound reproduction and production—tapes, computer programs of music-making, digital delay, AutoTune—that *enhances* the non-musical speaking voice, transporting it into the realm of musical sound.

[62] Pamela Z, "A Tool Is a Tool," in *Women, Art and Technology*, ed. Judy Malloy, Leonardo (Cambridge, Mass.: MIT Press, 2003), 358.

Chapter 3: The Intersections of Musical Frameworks and Unaltered Speech

While the last chapter focused on ways in which composers alter speech recordings to suit musical frameworks, this chapter is concerned with compositional techniques used when the recorded speech is presented in a largely unaltered form. Recall that the speech-to-song illusion requires the exact repetition of spoken phrases, fairly short snippets of speech. Not all speech lends itself so readily to interpretation as musical sound. In the previous chapter, we heard several examples in which the exact repetition of a recording of speech encouraged the listener to attend to the phonological features of speech after attending to its meaning. Indeed, Diana Deutsch and her colleagues hypothesize that exact repetition of the spoken phrase is what causes that listener to attend to the "musical" features of the sound, such as pitch and rhythm, instead of merely listening to the phonetic content in order to understand what's being said, as the listener is likely to do on first hearing.[1]

But composers who use recorded speech as a compositional object do not always use segmentation, exact repetition, or the technological manipulation of the recording in order to make speech conform to a musical framework. This chapter will focus on the compositional strategies found in musical works in which the composer presents speech in a largely unaltered form. Using recorded speech in this way in a musical setting results in several conflicts between speech and music that the composer must confront. The recorded speech excerpts in the works discussed in this chapter are likely to (1) be longer in duration, (2) be more complex in their structure, and (3) express more complex ideas. Longer excerpts of speech like those considered here pose a problem: speech does not have a periodic structure, which becomes particularly

[1] Diana Deutsch, Trevor Henthorn, and Rachael Lapidis, "Illusory Transformation from Speech to Song," *The Journal of the Acoustical Society of America* 129, no. 4 (April 2011): 2251–52.

apparent in longer phrases (whereas an excerpt that fits into a measure or two is more readily interpretable as musically rhythmic). There is also, within longer excerpts, more time for the pitch and rhythm of speech in the recording to diverge from an established pitch (usually tonal) or metrical framework.

Concerning the speakers themselves, these longer excerpts present the speaker's utterances in a largely unaltered form, mediated only by the recording process. The speaker and the meaning she conveys can strike the listener as more subjective, more personal, and more human than that of the speakers in the last chapter, who seemed more like musical devices for the composer than fully formed subjects. In this chapter, I will consider several of the techniques that composers have used to confront this divergence in the organization of sound between speech and musical frameworks. These include: (1) the use of musical motives based on important words or ideas that recur within the speech excerpt, (2) the adjustment or adaptation of musical frameworks to suit the phonological features of the speech, and (3) the treatment of speech as aleatoric or dissonant with respect to musical frameworks.

This chapter and the preceding one lead us to consider what ties together these musical works, which belong to many disparate genres and artistic movements. I believe that these works are a culmination of a musical interest in the late 20th century in how audio technology can be used to change and enhance composers', musicians', and listeners' relationship to sound. I posit that the use of speech recordings in music is an aesthetic stance rather than any sort of unified compositional technique. I will conclude this chapter by considering the confluence of influences and milieus that inform this aesthetic stance and its fixation with the technologically mediated voice.

3.1. Motivic materials

Musical motives can be used as a means of crafting musical form, emphasizing important musical gestures and bringing them to structural prominence. Similarly, in verbal communication themes and ideas may be repeated for emphasis, although this repetition is unlikely to be as exact sonically as the musical repetition of motives. Words may recur, or synonyms for those words may be used to underscore important concepts. In this section, I will explore instances in which these spoken "motives" are used in a musical context. In instances like this, repetition can set up the basis for the listener's hearing of musical form, making connections within the speech that might not have been apparent or significant until highlighted through musical means. Setting out, developing, and then returning to musical motives provides composers with a way to structure larger spans of music.

To explain what a musical motive is and how it relates to musical form and the listener's experience of a piece of music, I draw from Arnold Schoenberg's definition. Schoenberg envisions the motive as a recurring musical gesture with a *memorable* shape or contour. Schoenberg's idea of a musical motive is not linked to exact pitch intervals and rhythms, but rather to similar (but not necessarily exact) melodic contours and rhythms. One of the defining features of motive for Schoenberg is coherence, the notion that a musical entity is identifiable by the listener as similar to other musical entities, allowing for differences in intervallic organization and for emphasis on rhythm and contour.[2] Schoenberg allows for a wide range of what can be classified as belonging to a particular motive: iterations of the motive can be *exact* repetitions, *modified* by small changes (e.g., changing the pitch contour but maintaining the rhythm), or

[2] Arnold Schoenberg, *Fundamentals of Musical Composition*, ed. Gerald Strang and Leonard Stein (London: Faber, 1999), 8. See also Lawrence M. Zbikowski, "Musical Coherence, Motive, and Categorization," *Music Perception: An Interdisciplinary Journal* 17, no. 1 (1999): 7.

developed when there are more drastic changes to various musical parameters (e.g., changes in rhythm by insertion, omission, or condensation; insertion of embellishments; metrical shifts or changes in meter; changes in formal function.)[3] Lawrence Zbikowski suggests that Schoenberg's rather loose conception of motive accords well with the way humans structure their understanding through categorization. When placing things (objects, ideas, etc.) into categories, humans group according to highest level whose members have similar and recognizable shapes. The motive, for instance, is made up of smaller parts (pitches, intervals, durations) but with respect to how it is comprehended it is considered a whole entity.[4] Aniruddh Patel likewise posits that listeners can identify "similarity without identity" with regards to motivic similarity graded by features like surface acoustic features (timbre and articulation), midlevel structural features (melodic contour) and abstract features (implied harmony).[5] All this to say, Schoenberg's approach to motive is listener-based, relying on the cognitive ability of the listener to make connections rather than prescribing a set of rules for what a motive should be. In what follows I will adopt the same approach to motives, basing my analysis on the features that strike me as related between musical gestures, based on my own listening experience.

Oftentimes, the accompanying spoken words will help make clear these connections between musical motives. We have already seen one example of this technique in Chapter 1. In "Ants in My Eyes Johnson w/Drums," the introduction «I'm Ants in My Eyes Johnson» occurs twice, once at the beginning (0:02–0:05) and once roughly halfway through the video (0:25–0:28).[6] (Refer to Figure 1.3 for an outline of the form of this track.) The same drum pattern

[3] Schoenberg, *Fundamentals of Musical Composition*, 9–13.
[4] Zbikowski, "Musical Coherence, Motive, and Categorization," 12–13.
[5] Aniruddh D. Patel, *Music, Language, and the Brain* (Oxford: Oxford University Press, 2008), 197.
[6] David Dockery, "Ants in My Eyes Johnson w/Drums," YouTube, July 26, 2017, https://www.youtube.com/watch?v=G8UT6nkSWT4.

underscores it both times. This gesture provides a link between the two utterances, a signal to the listener that they are the same phrase delivered in a similar manner despite being about 20 seconds apart.

If we take a closer look at a spectrogram of the recorded voice for the two utterances (Figure 3.1), we can see how they might reasonably be grouped together based on their delivery. Melodically, both utterances follow the same basic contour: an up-down arch over «I'm Ants in My Eyes», a shift up for «John-» and a final fall for «-son». The only notable difference is that «I'm» in the second utterance has a more pronounced upward inflection. The timing of the syllables is also remarkably similar; «Johnson» falls just a hair sooner in the second utterance. This is perhaps due to the more frantic delivery of the second section of the monologue. Whereas the topic of the monologue in the first section focuses on what appliances Ants in My Eyes Johnson has for sale («I'm Ants in My Eyes Johnson here at Ants in My Eyes Johnson's Electronics»), the second section shifts to his condition («I'm Ants in My Eyes Johnson. Everything's black. I can't see a thing.»).

Figure 3.1. Spectrograms of spoken phrase «I'm Ants in My Eyes Johnson», "Ants in My Eyes Johnson w/Drums"

a. Voice only, "Ants in My Eyes Johnson w/Drums," 0:02–0:05

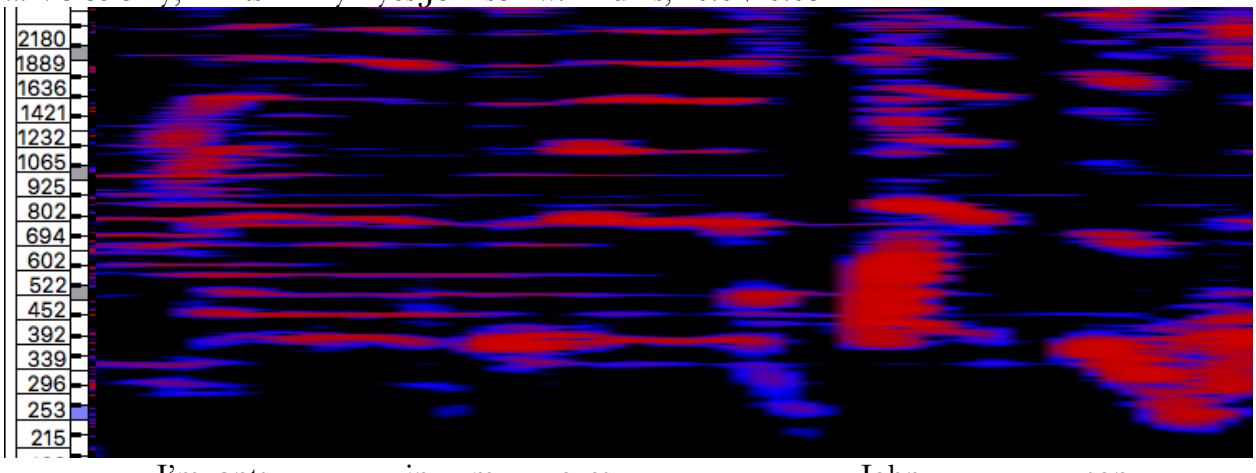
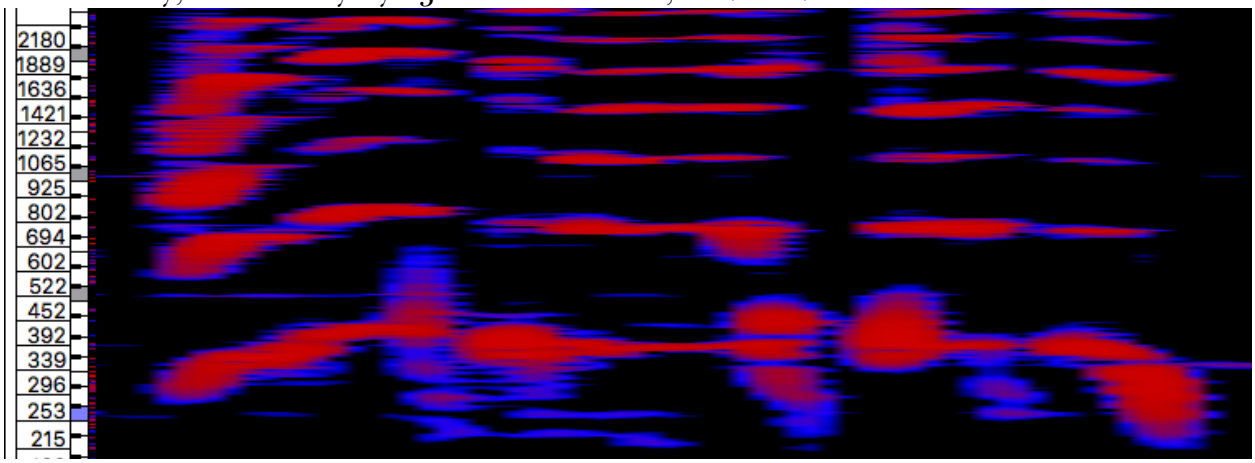


Figure 3.1, continued. Spectrograms of spoken phrase «I'm Ants in My Eyes Johnson», "Ants in My Eyes Johnson w/Drums"

b. Voice only, "Ants in My Eyes Johnson w/Drums," 0:25–0:28



This is a relatively simple example. There is only one spoken motive that is musicalized, and it is heard only twice. By underscoring these similarities in the voice with a recurring pattern in the drums, the drum motive suggests to the listener that this is the most important utterance in the monologue, both as the central theme and the only repeated phrase. It also highlights Ants in My Eyes Johnson's shift in focus from a sales pitch in the first section to a description of his ailments in the second section. This repeated motivic gesture helps divide the form of the musical work into two large sections, marking the beginning of each section. Indeed, the emergence of musical form from musical motives based on speech is perhaps the primary function of these motives within this repertoire.

We also saw an example of this—on a considerably more localized scale in the formal hierarchy—in Pamela Z's *Pop Titles "You,"* in which the list of pop song titles beginning with the word "you" is read as one unbroken excerpt of recorded speech.[7] While my analysis in Chapter 2

[7] A recording can be found on Pamela Z, *A Delay Is Better than a Disaster*, CD, ST-214 (Starkland, 2004). You may note that *Pop Titles "You"* has now come up as an example of both how a

focused on how the exact repetition of «you» sets up a basic metrical framework upon which the spoken phrases can be interpreted, I turn my focus here to the spoken phrases themselves. Each time «you» is heard as the subject of a linguistic phrase, it is interpreted by the listener as the beginning of that phrase (in both the linguistic and musical sense of the word). This results in the listener hearing these «you»s as a crucial moment in the formal structure of the music, with each «you» falling somewhere within the first beat of the musical phrase. *Pop Titles "You"* has limited linguistic content, and the word repetition sets up a sort of non-exact musical repetition—a motive based on a distinct phrase structure. This forms the basis for large-scale rhythmic expectations—that «you» will always be the beginning of a new phrase—resulting in metric dissonance when a spoken phrase is longer than the expected length of one measure. The «you»-as-subject motive serves to set up the listener's hearing of musical form by delineating phrases, structuring these musical phrases according to a segmentation that the listener makes based on linguistic content.

In order to dive deeper into how composers can use musical motives to organize the musical form of a spoken monologue, I turn now to a more complex example from Charles Spearin's 2009 experimental jazz album *The Happiness Project*.[8] Called "part spoken-word collection, part sociological study" by Pitchfork reviewer Rebecca Raber, this album is the result

composer can manipulate a speech recording to sound musical via looping and how a composer can apply musical structures to unedited speech recordings. This piece contains both the looped «you» and a long, unedited monologue in which Z reads the list of pop song titles. I am ultimately not trying to create a theory in which all examples fall neatly into one or another compositional strategy for musicalizing speech; I am instead examining ways in which different composers with a range of styles and backgrounds have approached the phenomenon of using speech as a compositional object. As a result, some pieces will draw on more than one of the compositional approaches I have outlined.

[8] Charles Spearin, *The Happiness Project*, CD, A&C039 (Toronto: Arts & Crafts, 2009).

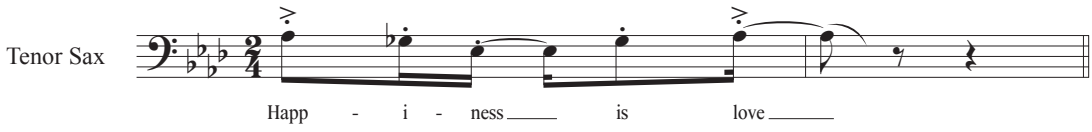
of Spearin's recorded interviews on the nature of happiness with his neighbors in Toronto.[9] Spearin used the recorded responses as the compositional basis for pitch and rhythmic frameworks. The album opens with "Mrs. Morris," a piece that sets the stage for the album both conceptually and musically. The basic compositional premise is the musicalization of the eponymous speaker's words by the saxophone. This musicalization assigns a musical pitch and rhythm—based roughly on the pitch and rhythm of the voice—to each syllable of her utterance (much as did the lead guitar and saxophone in "The Rick and Morty Band | Ants in My Eyes Johnson Cover on Drums, Guitar, Bass, and Sax" discussed in Chapter 1).[10] On the infrequent occasions that the interviewer speaks up, his voice is musicalized by the bass. The bass provides a fittingly muted timbre to accompany the somewhat muffled words of the interviewer, a marked contrast to the bite and clarity of the saxophone/main speaker combination. While many of the other tracks on this album rely on repetitions of short spoken phrases to create a groove, this particular track is through-composed, based on a monologue that sounds as though it has not been edited (aside from the repetition of the very first phrase). How then does Spearin create a sense of coherence and unity across the track when almost none of the spoken phrases are repeated directly? The deployment (and, importantly, development) of a compositional strategy focused around musical motives serves as the basis here for hearing form and making connections between spoken and musical phrases.

"Mrs. Morris" demonstrates the applicability of Schoenberg's broad approach to musical motive—particularly with regard to the listener's ability to hear and group together different-but-

[9] Rebecca Raber, "Charles Spearin: The Happiness Project," Pitchfork, June 1, 2009, https://pitchfork.com/reviews/albums/13089-the-happiness-project/.
[10] Sancheneering, "The Rick and Morty Band | Ants in My Eyes Johnson Cover on Drums, Guitar, Bass, and Sax," YouTube, September 7, 2017, https://www.youtube.com/watch?v=Pvidk3w4MfA.

similar musical ideas. Similar to the categories that humans use to organize their understanding of the world, a category that includes different forms of a motive will show typicality effects, with some members of the category being more typical than others.[11] Throughout "Mrs. Morris," Spearin uses modifications or developments of a single motive as an aural through line. The most typical version of the motive in "Mrs. Morris" would include (1) a descending contour, (2) a large falling leap between the penultimate and final pitches of the phrase, and (3) a roughly two-measure duration. As we will see, the interaction between the speech phrase and the musical setting plays with these expectations in different ways, creating different formal and affective results.

Figure 3.2. Introductory gesture (in concert pitch), Charles Spearin, "Mrs. Morris," 0:04–0:05



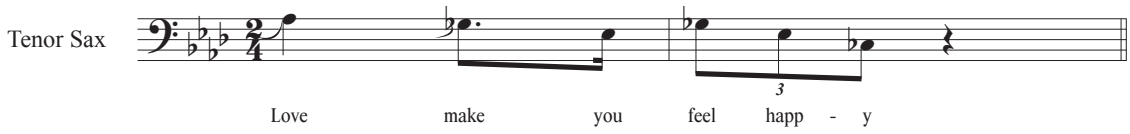
The track begins with an introductory gesture that, while it does not contain the motive that will be so integral to the structure of the piece, nevertheless sets up the basic groundwork on which the track is built. First we hear a musical phrase from the saxophone solo (sans voice), and then a spoken phrase from Mrs. Morris. Next, both the speaker and the saxophone repeat their phrases simultaneously, and at this convergence it becomes clear to me as a listener that the two gestures are intertwined and complementary (Figure 3.2). Although this initial gesture has a descending-then-ascending contour, the fall at the end of the final pitch serves as a sort of primer for the falling motive that will become prominent later in the piece. It is at this moment, the first simultaneous sounding of the voice and saxophone, that the panning of the track changes from

[11] Zbikowski, "Musical Coherence, Motive, and Categorization," 14–16.

mono to stereo: the saxophone in the left speaker and the voice in the right. This stereo panning continues throughout the track, providing a subtle spatial separation of the two sound sources that are ostensibly in "unison."

The main falling motive is next introduced accompanying the spoken phrase «love make you feel happy» at 0:06–0:08 of the track (shown in Figure 3.3). The main feature of the motive is its contour: so often in this piece the musical phrase (which corresponds with the spoken phrase) descends in pitch over its course, dropping off at the end to its lowest point on the final note.[12] Notice that this instance begins on the tonic of A-flat3 and over its course descends a sixth to C-flat3, with only one instance of upward motion. Notice also that the final interval, from E-flat3 to C-flat3, is the largest intervallic motion of the phrase. Both this contour (overall descending, but with occasional upward motion) and this emphasis on the final descent (by way of a larger interval) will be a common feature of this ever-changing but still recognizable motive.

Figure 3.3. Main motive, Charles Spearin, "Mrs. Morris," 0:06–0:08



[12] By now you may have noticed that contours of spoken phrases examined in this dissertation commonly end with descents (as in "Ants in My Eyes Johnson w/Drums," *Pop Titles "You"*, and the *Unbreakable Kimmy Schmidt* theme song). Indeed, the linguistic literature supports that this declination patter is a common feature of English spoken phrases. See Ivan Chow and Steven Brown, "A Musical Approach to Speech Melody," *Frontiers in Psychology* 9 (March 2018), 10–12; Jiahong Yuan and Mark Liberman, "F_0 Declination in English and Mandarin Broadcast News Speech," *Speech Communication* 65 (November 2014): 67–74; Philip Lieberman et al., "Measures of the Sentence Intonation of Read and Spontaneous Speech in American English," *The Journal of the Acoustical Society of America* 77, no. 2 (February 1985): 649–57. What makes the contour in "Mrs. Morris" notable, however, is the sharp drop-off on the final syllable.

As the track progresses, the falling motive is developed. Several musical parameters and gestures are at work in this development: range, formal function, phrase length, and intervallic content. The phrases become longer later in the track. Take for example the phrase heard at 0:18–0:23. As occurs throughout this piece, the musical phrase corresponds to the spoken phrase. In this case, «It make you feel happy to know that you're—you're my neighbor» is much more complex than «Love make you feel happy» or «Happiness is love», with the result that the musical phrase becomes more complex too. Despite its notably longer duration I nonetheless hear this phrase (shown in Figure 3.4) as a sort of expansion of the falling motive. The melodic contour follows a similar pattern: an overall falling gesture (despite some backtracking by minor or major second) with a sharp drop to C-flat3 in the final interval (here an even large interval of a fifth). Because of these similarities in contour and final pitch, I can conceptually group this longer phrase as a sort of modification of the main motive.

Figure 3.4. Expanded falling motive, Charles Spearin, "Mrs. Morris," 0:18–0:23

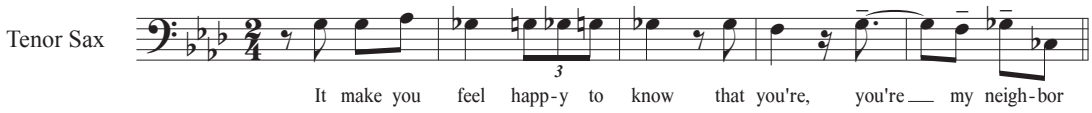
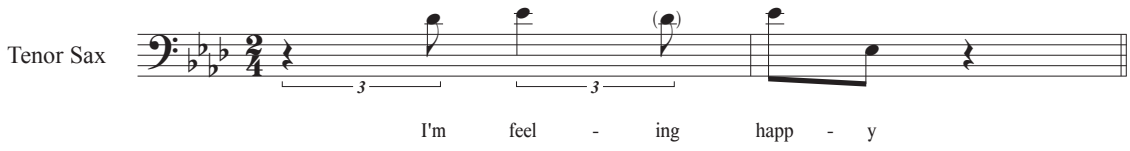


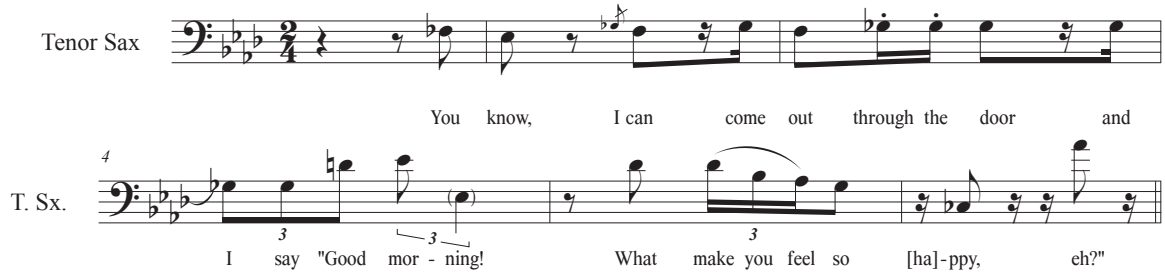
Figure 3.5. Development of falling motive, Charles Spearin, "Mrs. Morris," 0:36–0:37



At 0:36–0:37, Mrs. Morris says «I'm feeling happy». The corresponding musical phrase, because of its brevity and contour, pushes at the bounds of the motive (Figure 3.5). While the contour descends overall, there is very little melodic variation, just an oscillation between D-flat4

and E-flat4 concluded by a final drop off to E-flat3 an octave below. Its final octave drop-off is steeper than we've heard in previous examples of the motive. It is also in a markedly different pitch range—as Mrs. Morris's delivery becomes more animated over the course of her interview, and her voice rises in pitch space. Nevertheless, I still hear this phrase as part of the main motive. I believe that here the words have a great deal to do with my categorization. We've heard Mrs. Morris use the words «feel happy» twice previously: «love make you feel happy» at 0:06–0:08 and «it make you feel happy to know that you're—you're my neighbor» at 0:18–0:23. We might note that in this iteration, «feel(ing) happy» does not serve a concluding function as those other two examples, instead forming part of the neighbor-note oscillation. The phrase it's attached to, however, does conclude with a clear articulation of large descending interval, which is crucial to the motive's contour.

Figure 3.6. Development of falling motive, Charles Spearin, "Mrs. Morris," 0:45–0:52

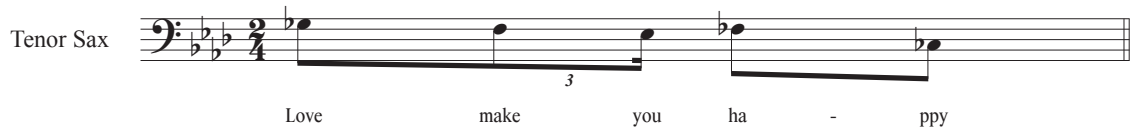


As the track progresses, the motive drifts further from its original shape in order to adapt to new formal functions. An example of this alteration in service to formal progress occurs with the spoken phrase «You know, I can come out through the door and say 'Good morning! What make you feel so happy, eh?'» at 0:45–0:52. This is quite a long phrase (shown in Figure 3.6), one that develops the motive in new ways. The motive occurs twice: I hear the first four measures as one iteration and the final two as another. The first iteration is an expanded near-inversion of the

motive—instead of descending, the phrase rises until it reaches the sharp drop-off that's so characteristic of the motive from E-flat4 to E-flat3 in the fourth measure. The second iteration follows a neat descending contour whose motion ends with a leap of an augmented fifth from G3 to C-flat4. In a sort of inverse parody of the final drop-off gesture, the final interval of this phrase is a huge leap *up* from C-flat4 to A-flat5. This upward leap is larger than any of the characteristic drop-off gestures, and serves to expand the pitch range. The moment occupied by this leap is a sort of midway point through the piece, so the expanded pitch range ratchets up the musical drama, pulling us farther from where we began both in terms of pitch space and in the structure of the motive.

The final iteration of the motive comes in the final phrase of the track: «love make you happy» at 1:23–1:24 (Figure 3.7). This iteration does a lot to bring the track full circle to its close. Although it is not an exact repetition of the first statement of the motive, it bears many similarities which make it sound like a clear modification of that motive. Perhaps most importantly, it returns the motive to the register in which it started. Despite beginning off-tonic on flat-$\hat{7}$ (G-flat3) instead of the tonic A-flat3, it ends on the same pitch as the first statement, C-flat3. It follows the overall descending contour with a momentary backtrack and concludes with a falling leap. Furthermore, it is lyrically quite close to the original, omitting only the word «feel». All of this makes this slightly compressed iteration of the motive feel like a reasonably conclusive final musical gesture that neatly ties the track together.

Figure 3.7. Final iteration of falling motive, Charles Spearin, "Mrs. Morris," 1:23–1:24



The use of musical motives in this repertoire of music works that make use of recorded speech brings to the fore patterns or similarities in speech that might not have been apparent when merely hearing prosody. Furthermore, the motives provide a sort of coherence that is essential to a *musical* hearing of form. According to Schoenberg, logic and coherence are necessary factors in identifying a comprehensible form.[13] In this section, we have seen examples of musical works that feature constantly shifting meters ("Ants in My Eyes Johnson w/Drums" and "Mrs. Morris"), that lack clear musical pitch frameworks ("Ants in My Eyes Johnson w/Drums," and *Pop Titles "You"*), and that exhibit little in the way of exact repetition. The hearing of musical motives can help the listener in parsing the musical surface into formal sections at both the level of the phrase (*Pop Titles "You"* and "Mrs. Morris") and the large-scale section ("Ants in My Eyes Johnson w/Drums) in works built around unedited recordings of speech, in which expectations of musical meter and pitch are at times undermined by the more unpredictable sounds of speech.

3.2. Adaptation of musical frameworks

In some musical works, as in the "Ants in My Eyes Johnson" example, composers and musicians attempt to suit their musical frameworks to the preexisting acoustical features of speech. The musical result of this molding is typically shifting meters and modulating pitch structures. I see adaptation as a composer's conscious choice to use a particular framework, one intended to best fit the recorded speech object.

To demonstrate how a composer might adapt the temporal frameworks of music to suit longer excerpts of recorded speech, I turn now to the movement "Arnold Schoenberg" from Peter Ablinger's *Voices and Piano* (1998–present). *Voices and Piano* uses complete, largely unedited

[13] Schoenberg, *Fundamentals of Musical Composition*, 1.

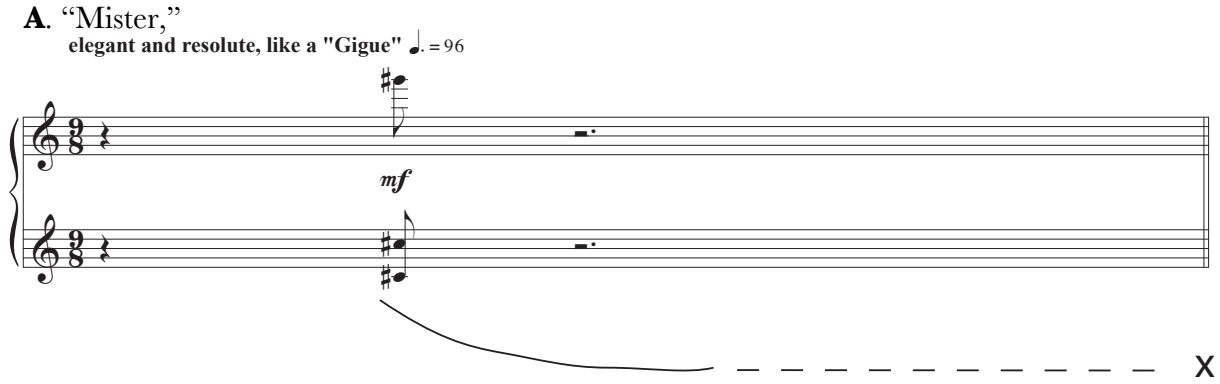
monologues as its compositional object and point of departure, focusing on one speaker per movement. This ongoing song cycle is scored for recorded speech and live piano. The recorded speech plays the role that the singer would in a traditional song cycle, serving as the focus, while the live piano performs an accompanimental role. Ablinger's piano accompaniment tends be composed in such a way that it places emphasis on each speaker's unique manner of speaking, which results in it rarely falling into one system of pitch or rhythmic organization for long. The movements vary wildly in style and character, using different compositional approaches to evoke the speakers' individuality. Ablinger uses the piano accompaniment in "Arnold Schoenberg" to emphasize the composer's somewhat halting English. Ablinger draws on the acoustic content of Schoenberg's speech—particularly the onset of syllables—as the basis for his musical metrical framework, a notated steady 9/8 meter. Despite this notated framework, Ablinger's reliance on speech as the basis for his rhythmic content often results in moments that seem hard to track within a metrical framework.

The source recording for "Arnold Schoenberg" is a voice memo to Schoenberg's publisher—in English—in which the composer lambasts his publisher for releasing a recording of *Ode to Napoleon* performed by a female singer. Ablinger uses an unabridged version of this voice memo, complete with pauses between phrases, as non-native English speaker Schoenberg searches for words. We can think of Schoenberg's monologue as "through-composed" prose, which does not repeat any phrases and does not fall into any obvious patterns. As a result, the musical phrases of Ablinger's composition, which are shaped by linguistic phrases, are of quite variable lengths.

The movement begins with a series of spoken phrases of varying durations that are accompanied by piano (Figure 3.8). Phrase **A** is merely an interjection of a single chord corresponding to the onset of the word «Mister». Phrase **B** introduces a longer spoken phrase

«you, in spite of my protests» underscored by chords of similar durations, forming the basis for hearing a very fast musical pulse. Phrase **C** reinforces the pulse set up in phrase **B** while introducing a variety of new rhythms, including longer durations that form the basis of a slower tactus. Phrase **D** is the most complicated yet, accompanying a longer spoken phrase («Leibowitz's performance of my *Ode to Napoleon* with a woman's voice») and presenting musical rhythms that are dissonant against the sense of meter that has emerged in phrases **B** and **C**. The musical meter of this excerpt is not immediately apparent or inherent to the speech rhythms, but rather develops over the course of the spoken sentence through the use of accentual cues that show metric potential.[14]

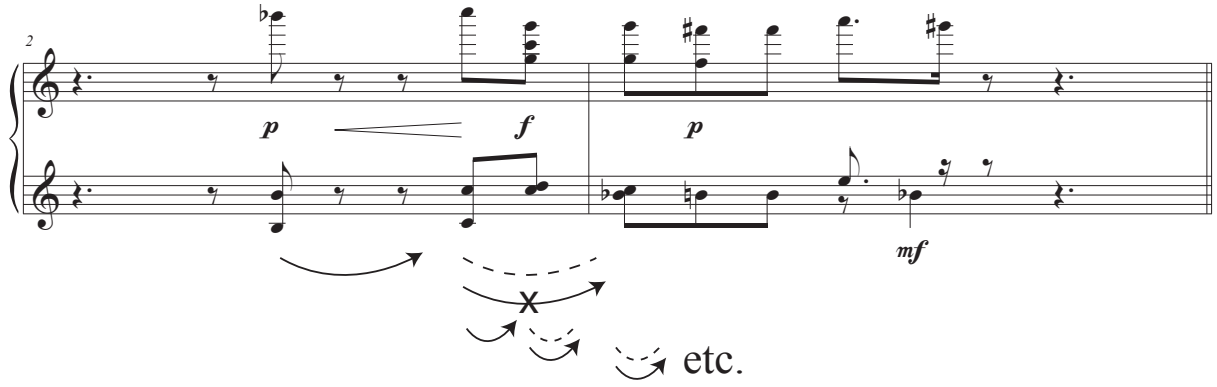
Figure 3.8. Varying phrase lengths and the emergence of a pulse in mm. 1–9, Peter Ablinger, "Arnold Schoenberg," *Voices and Piano*, 0:00–0:17



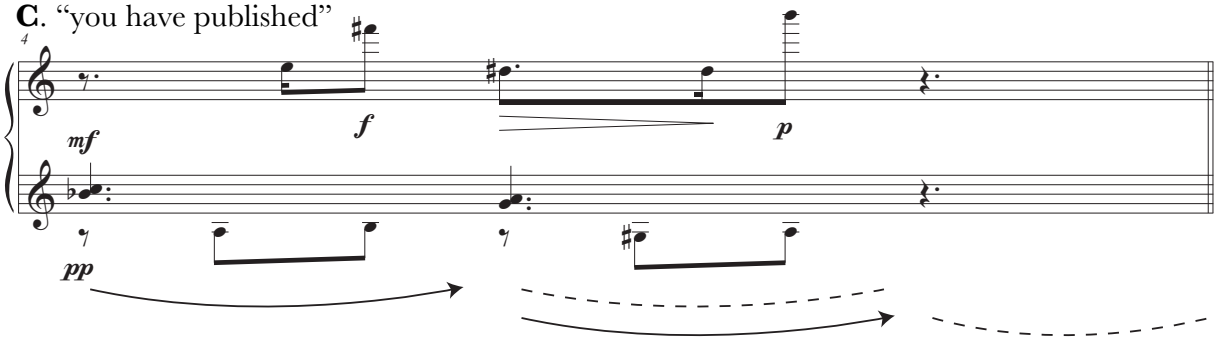
[14] For more on how accentual cues can show metric potential see Nancy Murphy, "The Times Are A-Changin': Metric Flexibility and Text Expression in 1960s and 1970s Singer-Songwriter Music," *Music Theory Spectrum* (forthcoming), which uses a process-based analysis of meter in order to explore moments of flexible timing in 1960s and 70s singer-songwriter repertoire.

Figure 3.8, continued. Varying phrase lengths and the emergence of a pulse in mm. 1–9, Peter Ablinger, "Arnold Schoenberg," *Voices and Piano*, 0:00–0:17

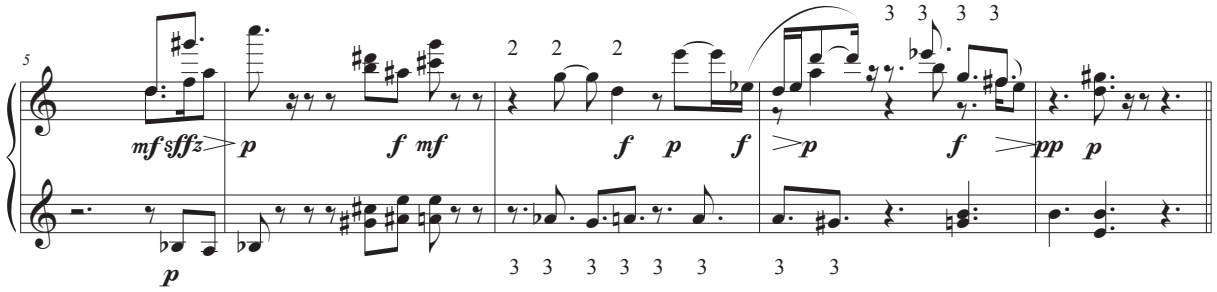
B. "you, in spite of my protests"



C. "you have published"



D. "Leibowitz's performance of my *Ode to Napoleon* with a woman's voice"



One way to understand how a listener might develop a metrical interpretation of the movement is to draw on Christopher Hasty's notion of projection. Hasty views meter as an emergent, relational property rather than a fixed network of parts or relations, which is an apt way of conceptualizing how Ablinger slowly builds the metrical framework across the first several

phrases in this movement.[15] Hasty proposes that "a mensurally determinate duration provides a definite durational potential for the beginning of an immediately successive event," and on the basis of this a listener may anticipate (or "project") a repetition of that same duration. If the first duration is followed by a second duration of a similar length, then the durational potential is realized and a sense of meter will arise.[16]

A projection-based analysis of the opening of the movement suggests that the process of setting up a recognizable meter is a slow one, occurring gradually over several phrases (labeled **A**–**D** in Figure 3.8). Following a quick interjection in phrase **A**, it is in the phrase **B** in mm. 2–3, with the repeated eighth notes at around 0:04, that a regularly-occurring pulse begins to become salient. We may here this as the beginning of the process of metrical emergence, where, according to Gretchen Horlacher, "metrical cues invite the listener to count, but their lack of consistency prohibits the confirmation of a meter at one or more levels."[17] Without any sense of accentuation of the beats in this phrase it is difficult to determine any deeper metrical framework: we only know that this regularly-repeated duration of around 288 bpm forms one level of periodicity.[18] It is only in phrase **C** (in m. 4 of the score and at 0:06 on the recording), however, that I begin to feel the actual tactus level of periodicity of 96 bpm indicated by the metronome marking in the score. Hasty suggests that higher-level durations are difficult for the listener to hold on to, positing instead that the listener "can use the smaller, more highly determinate

---

[15] Christopher Francis Hasty, *Meter as Rhythm* (New York: Oxford University Press, 1997), 67.
[16] Ibid., 84–91.
[17] Gretchen Horlacher, "Multiple Meters and Metrical Processes in the Music of Steve Reich," *Intégral* 14/15 (2000/2001): 271.
[18] Justin London, *Hearing in Time: Psychological Aspects of Musical Meter* (New York: Oxford University Press, 2004), 17–18 requires that there be at least two levels of periodicity, the tactus level and a slower, higher-level ordering of beats into measures. He further notes that there is likely to be a third, quicker level of subdivisions of the beat.

durations to enhance our feelings of continuation as a realization of projected potential."[19] I hear the notated dotted quarter note beat as a higher-level duration, one that is secondary to the primacy of the eighth note and its projective potential. Here, coinciding with the spoken phrase «you have published», the familiar compound-meter figure of a dotted eighth-sixteenth-eighth sounds twice in succession, confirming the importance of the lower metrical level of eighth-note subdivision and setting up the dotted quarter note as the tactus. Although we now have two levels of periodicity informing the metrical framework, there are still many silences—brought on by the stilted manner of Schoenberg's speech—which thwart my attempts at uncovering larger rhythmic/metric patterns in the following phrase **D**, patterns that might allow me to feel a clear sense of higher-level organization of the tactus into measures.

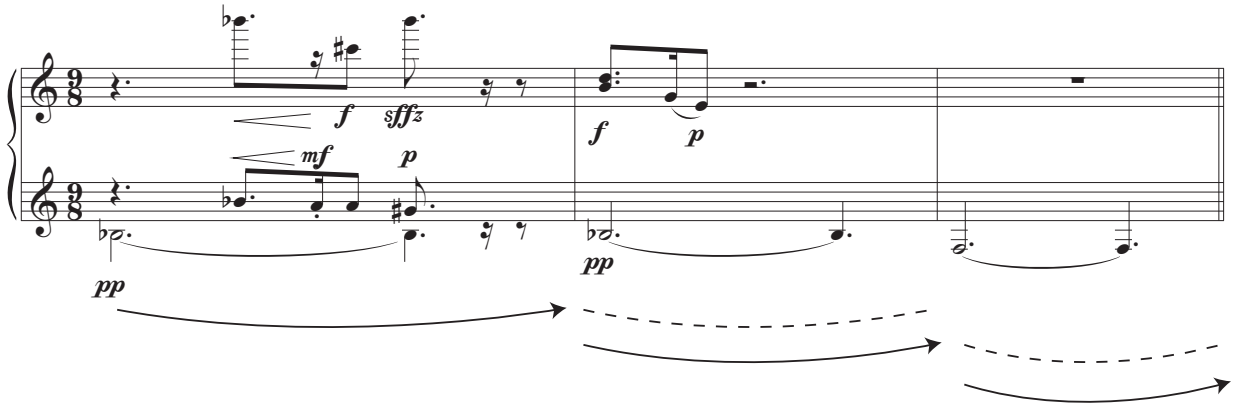
Another contributing factor to this difficulty in interpreting the musical meter in this phrase is Ablinger's use of notated "metric dissonance," which I put in scare quotes because I find it difficult to justify calling the effect dissonant without a clear concept of what consonance sounds like in this situation. Nonetheless, we can see groupings in this fourth phrase that fall outside of the expected divisions of the beat in 9/8. The most common is a dotted eighth note motif, a sort of G3/2 at the sixteenth-note level.[20] On the eighth-note level, we see a hemiola (G2/3) in m. 7. Lastly, because we have come to expect groups of three eighth notes beginning on the beats, the groupings of three eighth notes in mm. 5–6 have the potential to function as a sort of displacement dissonance—the association of layers of equivalent cardinality in a nonaligned manner—against the notated meter.[21] All of these potentials for metrical dissonance

[19] Hasty, *Meter as Rhythm*, 110.
[20] Recall here Krebs's discussion of grouping dissonance introduced in Chapter 2: Harald Krebs, *Fantasy Pieces: Metrical Dissonance in the Music of Robert Schumann* (New York: Oxford University Press, 1999), 31–33.
[21] Ibid., 33. I should note here that Hasty's theory projection has a problem in dealing with triple groupings. He goes to great lengths to justify them within his theory. He explains how the third

make this passage incredibly difficult to parse within the notated 9/8 framework, making it clear that the rhythmic structure of this phrase is governed more by Schoenberg's delivery of his speech than by any sort of traditional musical systems of organizing time. This metric ambiguity serves to musically reinforce the stop-and-start nature of Schoenberg's speech.[22]

Figure 3.9. The emergence of measure-level periodicity in mm. 10–12, Peter Ablinger, "Arnold Schoenberg," *Voices and Piano*, 0:18–0:23

E. "which I find terrible."



The use of measure-length pedal points beginning in m. 10 lays the groundwork for the higher measure-level periodicity. Phrase **E**, shown in Figure 3.9, presents the projections that

beat defers the projective potential of a two-beat grouping but also functions as a continuation of the second beat, creating a unique two-fold function of the third beat in a triple grouping. See Hasty, *Rhythm as Meter*, 130–139. I think that this convoluted explanation of a simple musical situation is a byproduct of an overly complicated theory, one that cannot unfold in real-time listening. Simply put, I don't think we as listeners make as many highly reasoned-through decisions when hearing meter. Western listeners have been acculturated to hearing triple groupings of beats through hundreds of years of music and can quite reasonably build metrical expectations based on hearing three like events instead of two.

[22] Justin London defines metric ambiguity as "a partial inhibition of our capacity for entrainment" and describes it metaphorically as "a social situation of conversational stumbles and interruptions" as it disrupts the potential for collective movement among listeners. See Justin London, "Cognitive and Aesthetic Aspects of Metrical Ambiguity," (Colloquium talk, University of Pennsylvania, Fall 2008). This metaphor is particularly apt in the case of Ablinger's "Arnold Schoenberg" because the piano accompaniment indeed serves as a sonic analog for conversational stumbles.

give rise to this periodicity. The dotted half note in m. 10 (0:18) sets up my expectation for a repetition of that durational event in m. 11, which is then realized and repeated again in m. 12. Thus, by the end of the fifth musical phrase—the end of the first complete sentence of Schoenberg's speech—Ablinger has built, layer by layer, a three-part metrical hierarchy. This moment coincides with the final clause of the sentence that Schoenberg speaks ("which I find terrible"). As the metrical hierarchy is unveiled over the course of mm. 1–12, so too is the reason behind Schoenberg's scathing voice memo to his publisher—the publication of what he views as an inappropriate performance of his composition.

Measure-length pedal points like the one found in Figure 3.9 recur somewhat frequently throughout the movement, usually in groups of two, three, or four. On an even higher level, although the movement as a whole is through-composed, these pedal points add motivic repetition that serves to unify its disparate elements. Given that they coincide with downbeats, they also serve to reinforce the notated meter. Because of their position at the end of Schoenberg's sentence, they serve a sort of closing function. When these pedal points appear later in the piece, they always fall at the beginning or end of a sentence, marking these important structural moments in the prose with a slowing of harmonic motion.

This movement highlights some of the conflicts between the temporal organization of speech and music, particularly bringing to the fore that, on its own, speech rhythm has no periodicity, no sense of temporal hierarchy. When the temporal organization of the music is tied so closely to speech rhythm, it is difficult to establish a clear tactus, let alone a fully fledged metrical hierarchy. The musical meter emerges gradually in this movement as Ablinger highlights similarities in syllable durations by assigning them eighth note rhythms as in Figure 3.8 and by using a natural pause in the speech to focus on the musical structure and establish measure-level periodicity in Figure 3.9.

*****

Another example of how composers adapt musical frameworks to the phonological features of recorded speech can be heard in the first movement, "Song," of James Tenney's *Song 'n' Dance for Harry Partch* (1999) for adapted viola, diamond marimba, strings, and percussion. Although this movement presents similar issues for the listener with regard to hearing meter—a symptom of structuring musical attacks around the onset of spoken syllables—what I will focus on in the following is how a composer can adapt pitch frameworks to suit recorded speech.

As with rhythm, there are fundamental differences between how pitch functions in speech and music. Musical pitch is usually a built around a stable set of pitch intervals. These fixed pitches are typically organized in a hierarchy where one pitch is the focal point (tonic) and other pitches are heard as stable or unstable in relationship to that central pitch.[23] Pitch in speech, on the other hand, slides up and down rapidly between tones, rarely dwelling on a specific tone for any perceptible amount of time. Speech melody is used to convey affective, syntactic, pragmatic, and emphatic information but—with the exception of tone languages like Mandarin and pitch-accented languages like Swedish and Japanese—does not influence word meaning.[24] Speech melody or intonation serves as a sort of icing on top of the linguistic cake, whereas musical melody is considered a primary parameter of music.[25] As I will show in the following analysis,

---

[23] For more on how listeners familiar with the Western musical idiom relation tones of the diatonic scale to one another, see David Huron, *Sweet Anticipation: Music and the Psychology of Expectation* (Cambridge, Mass.: MIT Press, 2006), 144–50.

[24] Aniruddh D. Patel, *Music, Language, and the Brain* (Oxford: Oxford University Press, 2008), 182. See Chapter 4 (pages 182–238) for a thorough literature review and comparison of speech melody and musical melody. For an approach to interpreting speech within a musical pitch framework (and its potential downfalls), see Chow and Brown, "A Musical Approach to Speech Melody."

[25] Leonard B. Meyer, *Style and Music: Theory, History, and Ideology* (Philadelphia: University of Pennsylvania Press, 1989), 14–16. This stance of speech melody may be an oversimplification. Shintel, Nusbaum, and Okrent argue that the acoustic dimensions of speech are a form of communication unto themselves. See Hadas Shintel, Howard C. Nusbaum, and Arika Okrent,

Tenney's system of pitch breaks with expectations regarding pitch frameworks in Western music in order to better accommodate the phonological aspects of speech.

In a similar fashion to the other pieces discussed so far in the this chapter, this movement is based around an unabridged recording of speech, in this case a reading of an excerpt from Harry Partch's treatise "Manual: On the Maintenance and Repair of—and the Musical and Attitudinal Techniques for—Some Putative Musical Instruments."[26] Tenney had a personal connection to Partch, having studied under him while at the University of Illinois in 1959–61, where he served as Partch's research assistant for six months.

Tenney's compositional process began by using Fast Fourier Transform (FFT) analysis (an algorithmic analysis that can translate complex sound signals into a grid of time and frequency) to quantify the pitch and temporal makeup of speech recordings. Tenney credits Randall Smith and Ernest Chokalis with using a signal processing program to analyze a recording of Tenney reading aloud text from Partch's essay.[27] He then used this information to render those sounds into notated musical structures. Although the speech recording forms the basis for the pitch and rhythmic content of the piece, it is not a part of the finished piece, which is orchestrated solely for acoustic musical instruments.

Tenney had long been interested in how speech could inform musical composition. In his master's thesis from 1961, written while at the University of Illinois studying under Partch, Tenney considers speech's influence on organizational structures in post-tonal music, particular

---

"Analog Acoustic Expression in Speech Communication," *Journal of Memory and Language* 55, no. 2 (August 1, 2006): 167–77.

[26] The "new (revised)" edition was originally written by Partch in 1963. It was first published in Harry Partch and Philip Blackburn, *Harry Partch*, Enclosure 3 (Saint Paul, MN: American Composers Forum, 1997), 321–362.

[27] James Tenney, *Song 'n' Dance for Harry Partch: For Adapted Viola, Diamond Marimba, Strings and Percussion* (Lebanon, NH: Frog Peak Music, 1999), iii.

in terms of meter and form. He notes that asymmetrical phrase structures are a common feature in 20th century composition, claiming that such structures engender "prose-like rhythmic development" and are derived from "desire to replace the measured simplicity of verse and dance-like rhythms with the freer rhythms of speech."[28] For Tenney it seems that the sound of speech is integral to contemporary music composition, and the organizational structure of "Song"—written three decades later—is a natural extension of this thinking.

Previous to writing *Song 'n' Dance for Harry Partch*, Tenney had experimented with speech as a compositional object in *Three Indigenous Songs* (1978) and *Ain't I a Woman?* (1992).[29] Robert Wannamker calls these three works Tenney's "speech-modeling pieces," each being an attempt at the instrumental synthesis of speech.[30] Tenney's approach to the musicalization of speech in all three pieces involves transcribing formant frequencies into pitched instrumental sounds and using percussion and pizzicato strings to mimic the articulation of consonants. Notably, in *Song 'n' Dance*

---

[28] James Tenney, "Meta ≠ Hodos: A Phenomenology of Twentieth-Century Musical Materials and an Approach to the Study of Form (1961)," in *From Scratch: Writings in Music Theory*, ed. Larry Polansky (Urbana: University of Illinois Press, 2015), 24. Indeed, these phrase lengths based on the rhythms of speech are a common feature of the music found in this chapter, having already been noted in "Mrs. Morris" and *Pop Titles "You"*.

[29] The compositional process for *Three Indigenous Songs* did not involve a speech recording. Instead, Tenney derived his pitch content for his chosen text from the average formant frequencies found in Grant Fairbanks and Patti Grubb, "A Psychophysical Investigation of Vowel Formants," *Journal of Speech and Hearing Research* 4, no. 3 (September 1961): 203–19. For *Ain't I a Woman?*, with text from Sojourner Truth's 1851 speech at the Woman's Rights Convention in Akron, Ohio, Tenney used FFT analysis to derive his pitch and rhythmic information for the instruments from a recording of the text done by Michele George. The piece also includes a live speaker. No commercially available recordings of this piece exist, and it has fallen out of performance in recent years due in part to its appropriation by a white male composer of a Black woman's speech (see Wendalyn Bartley, "Concert Report: The Fine Art of Social Consciousness – 'James Tenney: Resistance,'" *The WholeNote*, October 24, 2017, https://www.thewholenote.com/index.php/other-media/blog/concert-reviews/27382-concert-report-the-fine-art-of-social-consciousness-james-tenney-resistance).

[30] Robert A. Wannamaker, "The Spectral Music of James Tenney," *Contemporary Music Review* 27, no. 1 (February 1, 2008): 107–08.

Tenney employs a 72-tone microtonal system of pitch in order to more accurately reproduce the fundamental frequencies and harmonic partials of vowels.[31]

Song 'n' Dance is arguably Tenney's most nuanced attempt at what I call instrumental speech synthesis: the artificial production of human speech by non-human sources via instrumental, musical means. This variety of speech synthesis taps into not only the sounds of speech but also the style of the composer who wrote the words being spoken. Tenney makes use of two of the instruments that Partch designed, the adapted viola and the diamond marimba and—most remarkably for this analysis—a microtonal system modeled on Partch's compositions. It is worth noting that Tenney uses "Song" as an homage to Partch's compositional style by borrowing aspects of musical stylistic frameworks from Partch. I wish to focus here, however, on how Tenney's chosen pitch framework serves as an adaptation that better accommodates the phonological nuances speech.

In order to demonstrate what this means musically, I turn now to an excerpt from the first movement of *Song 'n' Dance for Harry Partch*, "Song." Figure 3.10 provides an excerpt of the score (mm. 1–5) that shows how Tenney sets the first spoken sentence.[32] The fundamental frequency of each vowel is represented by the adapted viola on the bottom line with the corresponding words from the recorded speech object transcribed beneath. This line functions as a kind of melody. The other strings play the formant frequencies as harmonies above this. At the top of the system are percussion parts that recreate the articulation of consonants. Between the percussion and the strings, Tenney has given a phonetic transcription of the spoken phrase with

[31] See author's notes in James Tenney, *Song 'n' Dance for Harry Partch: For Adapted Viola, Diamond Marimba, Strings and Percussion* (Lebanon, NH: Frog Peak Music, 1999), [iii].
[32] Ibid.

vowels on the upper line corresponding to the percussion and vowels on the lower line corresponding to the strings.

Figure 3.10. Score, mm. 1–6, James Tenney, "Song," *Song 'n' Dance for Harry Partch*

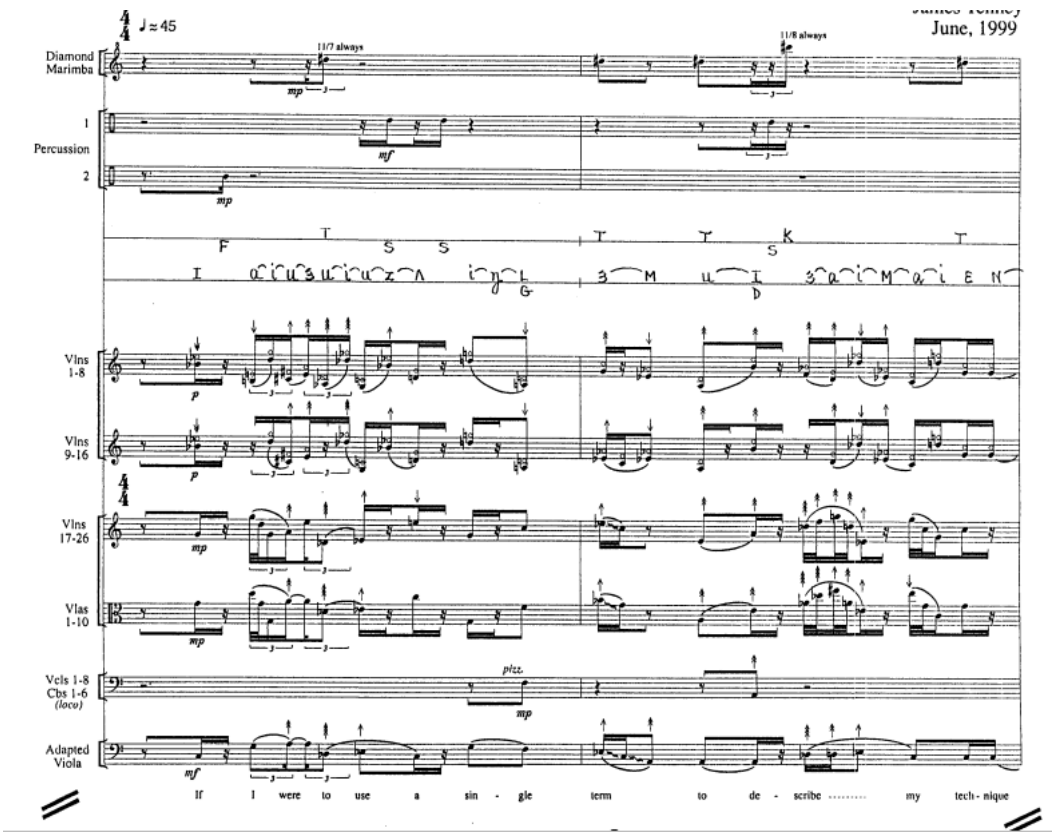
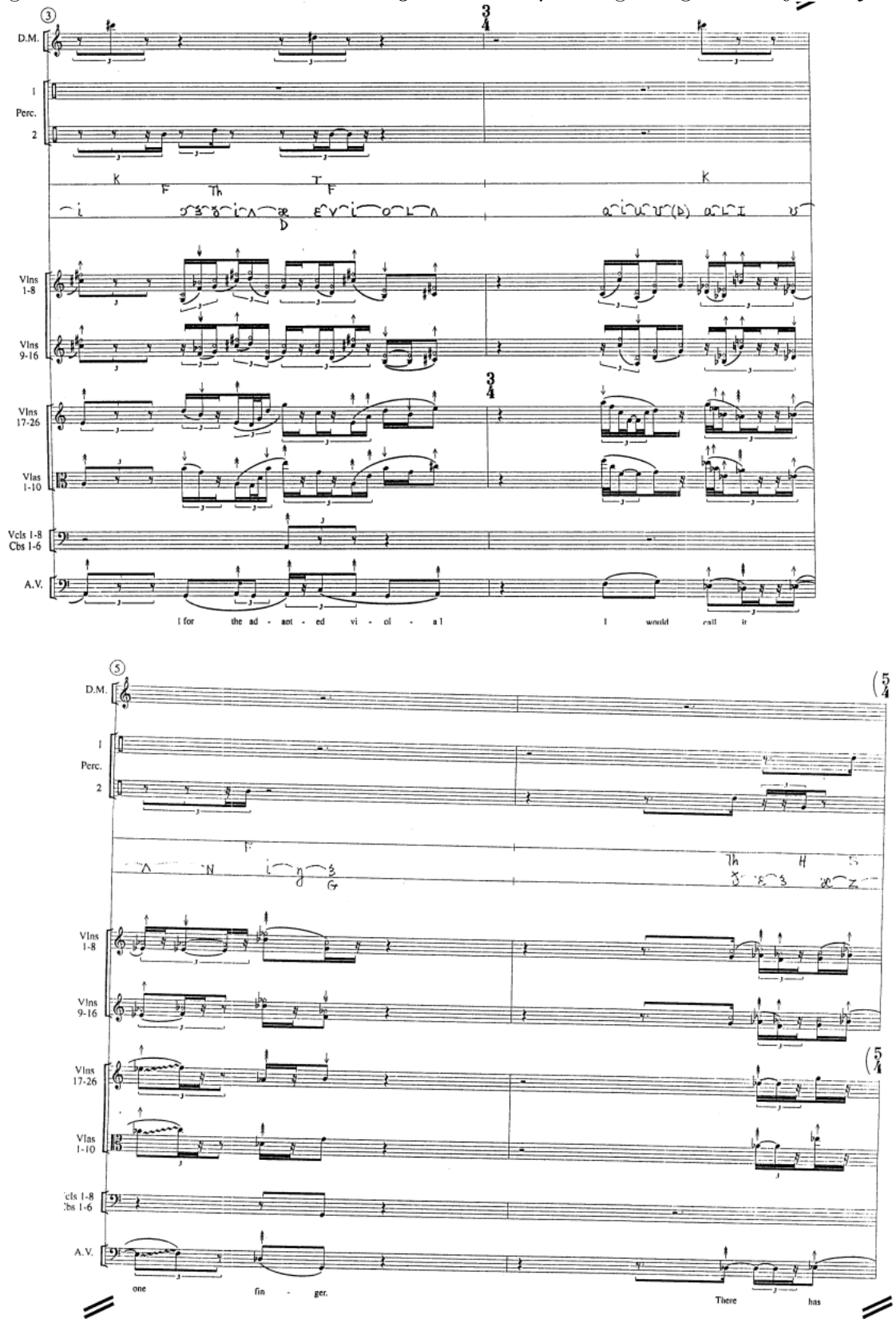


Figure 3.10 continued. Score, mm. 1–6, James Tenney, "Song," *Song 'n' Dance for Harry Partch*



We've seen previous examples in which the pitch content of speech was regularized into a musical pitch hierarchy with which the Western listener is likely familiar.[33] Tenney's approach to musical pitch organization, by contrast, takes its cues more directly from the pitches of speech in several ways. Most obviously, there's no clear pitch hierarchy. It is of course quite common for composers of Western art music in the 20th century to use post-tonal approaches to pitch organization, but what's remarkable here is how Tenney's compositional choices allow his instruments to more closely model the fluctuating nature of speech melody through microtonality and glissandi.

The 72-tone scale quite effectively disabuses the listener of any expectations of familiar pitch hierarchy. This finely detailed grid of musical pitch allows Tenney to more closely approximate the non-discrete pitches of speech. These microtonal pitch frameworks, which were employed by both Tenney and Partch, produce in the listener a feeling of "estrangement," a term coined by Alejandro Madrid to discuss listener and performer reception of 20th-century Mexican composer Julián Carillo's microtonal pianos. Estrangement is "a theoretical tool used to explore the aesthetic relation between [microtonality] and the mainstream music system from the common practice period." It results in a cognitive dissonance between expectation and what is heard that provides new and unexpected sonic experiences.[34] Here Tenney estranges the listener from the familiar 12-note system of pitch organization in order to distance the listener from a

[33] The example of «sometimes behave so strangely» found by Deutsch falls neatly into B major in her transcription. In Steve Reich's *Different Trains* from Chapter 2, the pitch content of speech is transcribed for the strings into various modal systems that shift for each spoken utterance, an example of how a composer may adjust pitch frameworks to suit the speaker. Charles Spearin's "Mrs. Morris" allows for pitches outside of the diatonic scale, but roots these with an A-flat tonal center and expands the available pitch resources by way of stylistically appropriate modal mixture, chromatic inflection, and scoops and falls between pitches.

[34] Alejandro L. Madrid, *In Search of Julián Carrillo and Sonido 13* (New York: Oxford University Press, 2015), 261.

traditional musical experience; by creating this distance, Tenney encourages his audience to listen for the phonological features of speech through the lens of music.

Tenney's use of glissando in "Song" is likewise an attempt to adapt musical pitch to approximate speech. When we speak of fundamentals and formants of vowels we are usually actually talking about the average of frequencies that fluctuate over the duration of the vocalization of the vowel. Tenney captures this fluctuation by notating multiple fundamentals per syllable («term» and «-scribe» in m. 2, «one» in m. 5) and notating glissandi for the viola and sometimes the other strings, which produces slides between pitches that approximate a smooth, continuous pitch fluctuation. "Song" therefore attempts not only to provide greater nuance in pitch differentiation but also to capture musically the variability of pitch across vowels in the human voice.

At a temporal level, Tenney's "Song" exhibits many of the same tensions between notated meter and heard meter we saw previously in Ablinger's "Arnold Schoenberg." On a pitch level, "Song" is structured according to the phonological features of speech but filtered through the resources provided by microtonal frameworks. What's important is that both time and pitch are interpreted musically. The speech rhythms are musically regularized against a tactus, even if at times that tactus is not clear to the listener. The pitches are set within a system of discrete pitches, albeit one that includes a large number of possible pitches, the wealth of which is further complicated by glissandos between these pitches. Nevertheless, this is still clearly meant to be interpreted by the listener as music and not as a direct sort of speech synthesis (an approach we will consider later in this chapter in Peter Ablinger's *Quadraturen*).

The kind of estrangement brought on by Tenney's unusual approach to musical pitch ultimately can inspire a transcendence of normativity: Tenney's musicalization of speech encourages the audience toward a different kind of listening, one that acknowledges musical

features but points toward new directions in the organization of sound along the lines of Tenney's aforementioned interest in compositions that are informed by the sonic structure of speech.

As both Tenney's "Song" and Ablinger's "Arnold Schoenberg" have shown, the adjustment of musical frameworks to suit speech provides a way to build music around excerpts of recorded speech that are much longer than the ones found in Chapter 2 and therefore less likely to fit neatly into traditional Western musical frameworks. By building a flexible system around the phonological features of speech, composers can more closely mimic the speech recording but at something of a loss to listeners' musical expectations. These works challenge the listener to hear beyond hierarchical meter and pitch and to approach the hearing of music with open ears.

3.3. Speech as dissonance or aleatory

In the preceding two sections we saw musical works in which the composers consciously translated the phonological features of relatively long speech recordings into musical frameworks. This final section will deal with pieces in which the composers acknowledge the aural differences between speech and music by purposefully not slotting speech sounds neatly into musical frameworks. I identify two means by which composers accomplish this. First, composers may position recordings of speech as dissonant against musical frameworks in order to play up the differences between speech and music. Second, we'll see a unique example in which the composer more or less forsakes traditional musical frameworks in favor of an approach to composition based on the phonological features of speech.

Returning to Peter Ablinger's *Voices and Piano*, the movement "Bonnie Barnett" uses rhythmic dissonance in order to explore the two contrasting ways music and speech display the quality of smoothness. This movement uses excerpts from vocalist Bonnie Barnett's radio

program Trilogy, which showcases new and avant garde music. As a trained vocalist, Barnett's speaking voice has a quite melodious quality, one that Ablinger emphasizes through his piano accompaniment.

The movement begins with just Barnett's recorded speaking voice and a monophonic bass line from the piano. Arranged much like the saxophone and voice in "Mrs. Morris," the piano here is meant to underscore the pitch of the voice. Unlike Spearin in "Mrs. Morris," however, Ablinger chooses to keep the piano moving in a steady quarter-note rhythm (that only rests when the voice does) rather than choosing rhythms that more directly suit the speech rhythm. This holds true even later in the movement, when the piano accompaniment branches out into a homophonic texture: the piano is always heard with a steady rhythmic beat.

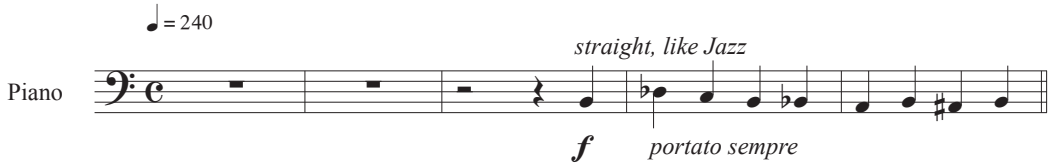
To demonstrate the effect of a steady musical rhythmic flow on our hearing of speech, consider the music of the first spoken phrase, «This is KRLU Los Angeles». Figure 3.11 shows (a) the score with the piano bass line and (b) a wave-form analysis taken from the Nicolas Hodges's recording of Ablinger's composition, with the onsets of syllables marked in green above and the note attacks from the piano marked in red below.[35] We can see from Figure 3.11b that the first five syllables and beat onsets are more or less aligned, setting up an expectation that beats and syllables will correspond with one another. As the phrase continues, it becomes clear that while the beats being played by the piano are roughly equal, certain syllables—particularly «L» and «-gel-»—are markedly shorter than average. With the onset of the first of these, «L», the speech rhythm diverges from the musical rhythm. The second of these, «-gel-», sets up a realignment of beats and syllables. The result is a sort of dizzying non-alignment with ten syllables in a span of

[35] Nicolas Hodges, *Voices and Piano*, CD (Vienna: Kairos Music, 2009).

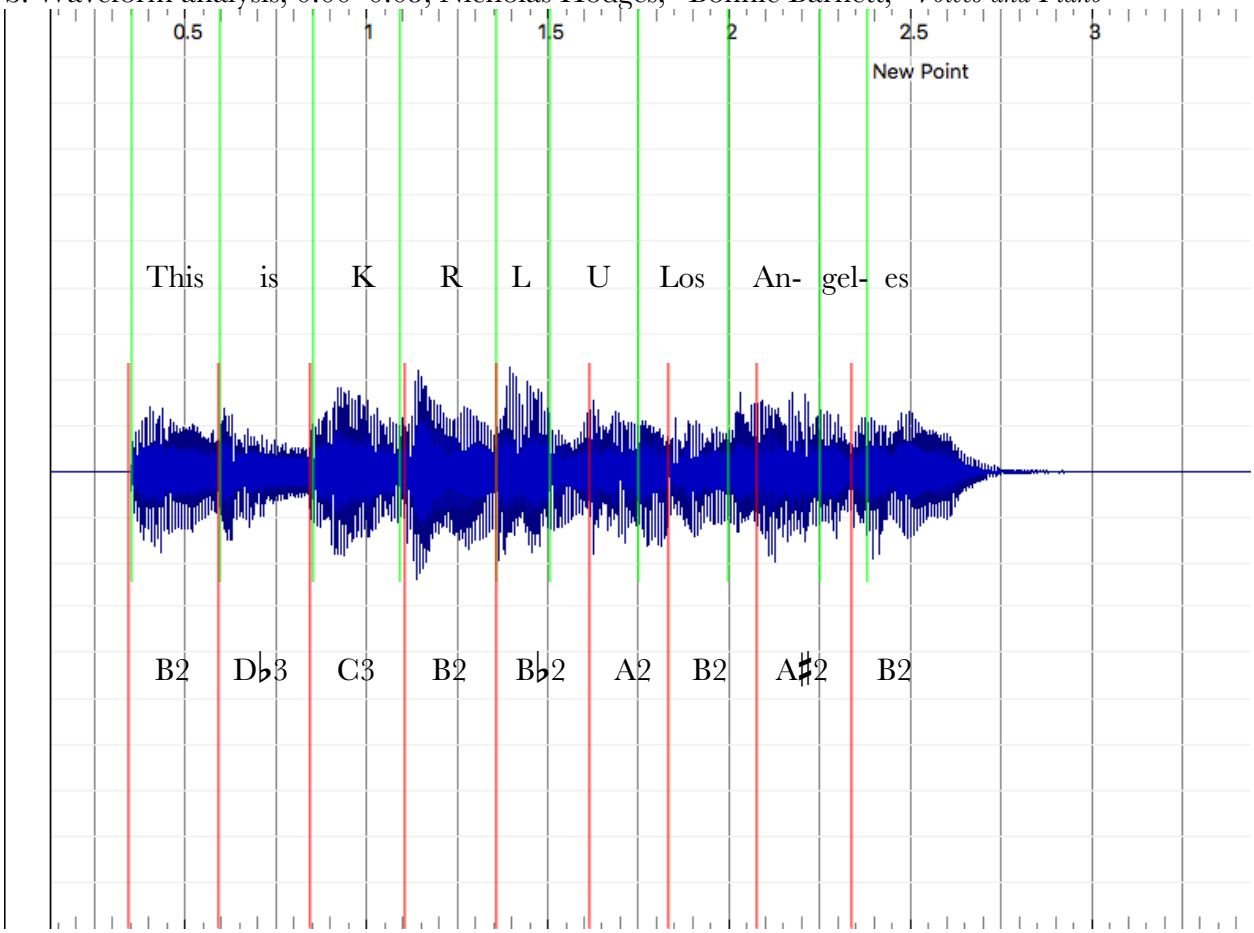
nine beats—close but just far enough off to be noticeable for a listener. The steady pulse of the piano provides a tactus-level framework against which the speech rhythm is heard as dissonant.

Figure 3.11. First phrase: «This is KRLU Los Angeles», Peter Ablinger, "Bonnie Barnett," *Voices and Piano*

a. Piano Score, mm. 1–5



b. Waveform analysis, 0:00–0:03, Nicholas Hodges, "Bonnie Barnett," *Voices and Piano*



There is a non-alignment between the rhythm of the voice and the piano, but it seems to me that Ablinger's choice of an even, repetitive musical rhythm is an attempt to mirror the

smoothness of Bonnie Barnett's voice through musical analogy, the reasoning process through which a listener can match the relational structure of music to a non-musical phenomenon.[36] This is a situation in which we can clearly see a difference between natural language and music in how expressiveness is communicated. In this case, smoothness is a metaphor applicable to both speech and music. Neither Barnett's voice nor the piano bass line has the physical, tactile property of being smooth, but rather something in their structure encourages a conceptual mapping from the domain of touch onto the domain of sound that helps us understand what we're hearing.[37] In the voice, the sound of smoothness is a connectedness from syllable to syllable. In the piano accompaniment, the sound of smoothness is a legato articulation (likewise a connectedness) and evenly spaced note attacks in the form of a steady quarter-note rhythm. Although this example demonstrates a disconnection between speech rhythm and musical rhythm that manifests as a dissonance, it also highlights a metaphorical likeness between the voice and piano.

*****

Composers may also place speech (and its "melody") against musical frameworks to create dissonances in pitch. To illustrate this, I will explore an excerpt from René Lussier's (b. 1957) experimental rock album *Le trésor de la langue* (1989).[38] *Le trésor de la langue* uses recordings of Québécois speakers to explore the position of the French language in Québec culture, with samples ranging from chatter on the streets of Montréal to news broadcasts to political speeches.

---

[36] For a more in-depth investigation into the cognitive processes at work in such mappings, see Lawrence M. Zbikowski, "Music, Analogy, and Metaphor," in *The Routledge Companion to Music Cognition*, ed. Richard Ashley and Renee Timmers (New York: Routledge, 2017), 501–512.
[37] For more on cross-domain mappings and metaphor, see Idem., *Conceptualizing Music: Cognitive Structure, Theory, and Analysis* (New York: Oxford University Press, 2002), 65–68.
[38] René Lussier, *Le trésor de la langue*, CD (Montréal: Ambiances Magnétiques, 1989).

The fourth track, "Vestibule/Çé ça qu'on va faire!/Le gars du Irving (direction 4)/Le tic-tac d'la veille/Le blues des résultats/Lend'main d'veille/Qui se souvient?," is divided into seven main sections, each centered around a different speaker or set of speakers.[39] As the speaker changes, so too does the mood, tonality, texture, and orchestration of the music, another example of the way musical frameworks are adapted to suit the speech. I will focus on the "Çé ça qu'on va faire!" section, which uses recordings from two speakers, in order to more closely examine how Lussier uses pitch dissonance to accommodate the expressive delivery of a monologue on top of a tonal, metrically regular musical framework.

Following a mostly instrumental build-up accompanied by intermittent radio static in the opening "Vestibule" section, a man's speaking voice (that of Québécois comedian Yvon Deschamp) enters at 0:44 accompanied by electric bass to kick off the "Çé ça qu'on va faire!" section. The bass "doubles" the rhythm and pitch of the speech, a texture we've heard before in "Mrs. Morris" and "The Rick and Morty Band | Ants in My Eyes Johnson Cover on Drums, Guitar, Bass, and Sax." The final line of Deschamp's monologue, «Çé ça qu'on va faire!» ["That's what we're going to do!"], uttered at 1:03, then becomes the basis for the groove that dominates a second, groove-based section, which is transcribed in Figure 3.12. This spoken phrase is musicalized through repetition, recurring as an anacrusis every two or four measures such that it emphasizes strong measures. It also serves to set up the tonic B-flat, which is reinforced by the bass line.

[39] For a deeper dive into the different ways in which Lussier uses recordings of speech in the music throughout the album, see Ana Dall'Ara-Majek, "Le trésor de la langue de René Lussier: Documentation d'un voyage, de la quête linguistique au débat politique," *Circuit: Musiques Contemporaines* 28, no. 3 (2018): 72–75.

Figure 3.12. Melodic and harmonic transcription, René Lussier, "Vestibule/Çé ça qu'on va faire!/Le gars du Irving (direction 4)/Le tic-tac d'la veille/Le blues des résultats/Lend'main d'veille/Qui se souvient?," 1:03–1:36

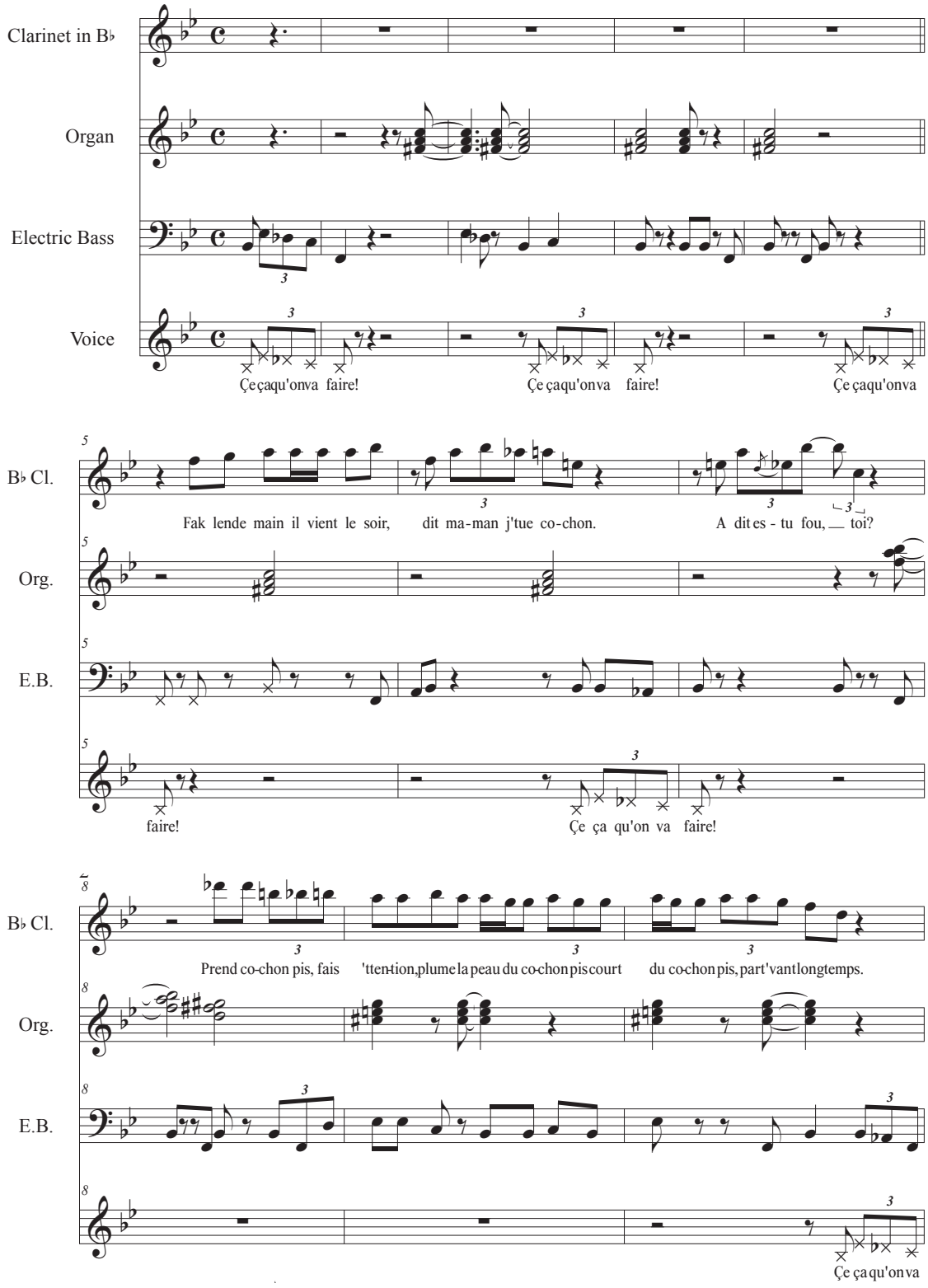
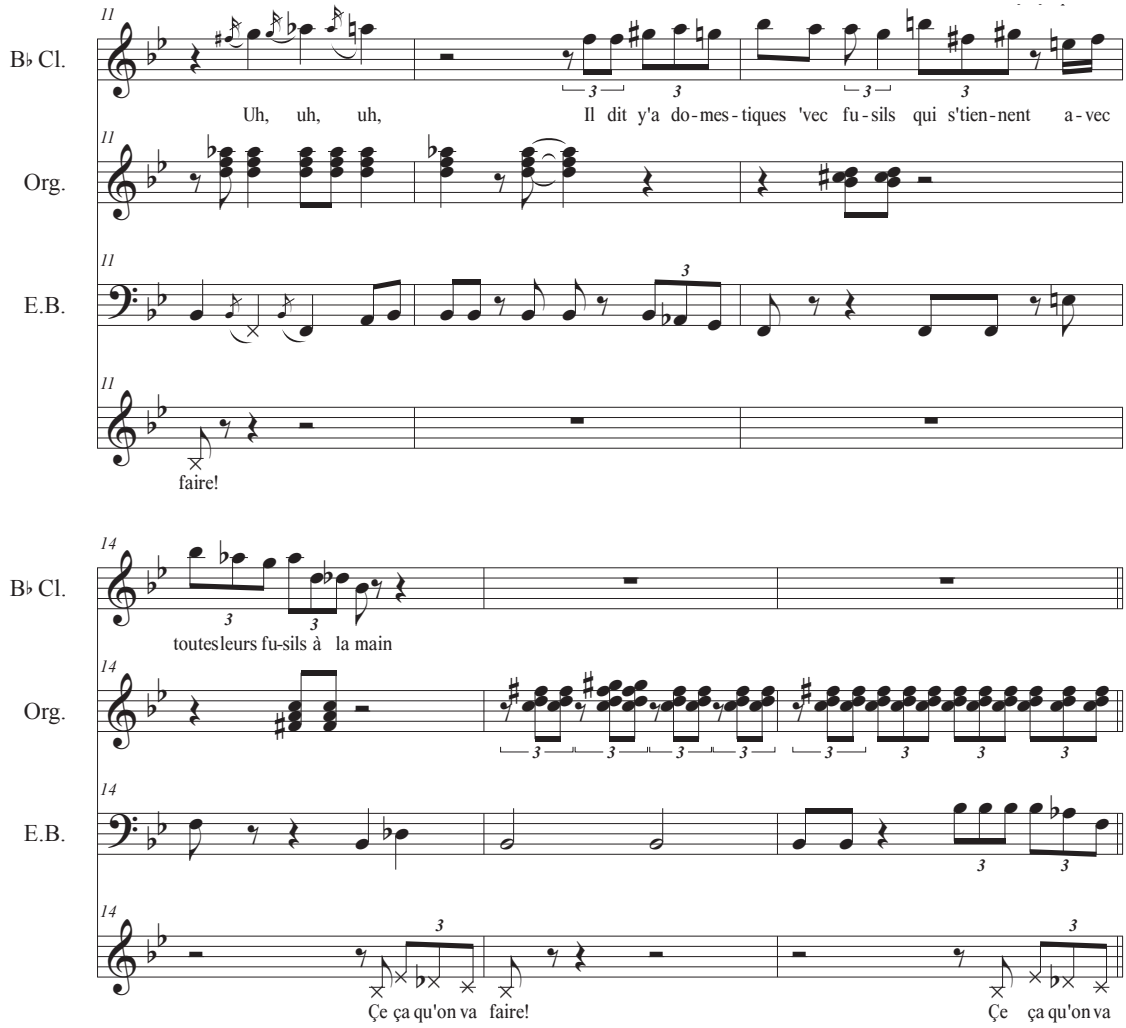


Figure 3.12, continued. Melodic and harmonic transcription, René Lussier, "Vestibule/Çé ça qu'on va faire!/Le gars du Irving (direction 4)/Le tic-tac d'la veille/Le blues des résultats/Lend'main d'veille/Qui se souvient?," 1:03–1:36



On top of the groove led by Deschamp's «Çé ça qu'on va faire!», a woman's voice tells a story, doubled by a clarinet melody. The clarinet's musical transcription of the woman's monologue does not, however, fit neatly into the temporal and pitch framework established for the «Çé ça qu'on va faire» groove.

In terms of the temporal intersections of speech and music, the drums, bass, and organ divide this section neatly into four-measure phrases. Following a four-measure introduction, mm.

5–16 of the transcription set out a 12-bar blues progression. The phrasing of the woman's voice/clarinet line, however, does not sit neatly within this form. Instead, the woman's voice/clarinet enters a beat late for its first phrase in m. 5 and two beats early for the second phrase that begins in m. 9 and the third phrase that begins in m. 13.

In terms of pitch, the clarinet melody often veers away from the key in order to more closely approximate the pitch content of the recorded speech. In mm. 6–8 and 12–14 in particular, the clarinet's pitch content seems to veer toward free atonality, no longer moored by a tonic pitch. Between the bass and the clarinet we essentially have two overlapping, conflicting pitch frameworks, each built to suit a particular speaker. The dissonant harmonies of the organ serve as a sort of link between these two worlds of pitch organization—they are rooted to the «Çé ça qu'on va faire» groove (especially in terms of rhythm, fitting into the 12-bar blues progression even if the pitches are off-kilter within that pitch framework) but open up the possibilities for pitch relationships beyond the tonal system centered around B-flat that is laid out by the bass. The clarinet is noticeably low in the mix and the recording of the woman's voice is poor and difficult to understand, positioning both as a secondary element to the overwhelming presence of «Çé ça qu'on va faire» and its accompanying groove. By pitting these two voices and their accompanying instruments against one another, Lussier underscores the differences between the two speakers, giving each their own soundworld.

Both "Bonnie Barnett" and "Çe ça qu'on va faire!" rely on clear musical frameworks that are juxtaposed with speech. In doing so they highlight the differences in pitch and rhythmic organization between music and speech, demonstrating how their systems of organization are in some way incompatible. I turn now to another of Peter Ablinger's works, *Quadraturen* (1997–2004), which is a study in pushing the boundaries of music and speech—an attempt at turning speech into music, but on the organizational terms of speech. Ablinger uses the piano to

reproduce the voice—another example of instrumental speech synthesis—recreating the phonological features of vocal utterances through instrumental means. In doing so, he forsakes traditional musical frameworks to create something that sounds aleatoric by musical standards.

Throughout *Quadraturen*, Ablinger uses instruments to produce a kind of musical, speech-synthesis-based analysis of an unedited speaking voice. I will focus on the movement "A Letter from Schoenberg," from *Quadraturen III*, because here Ablinger uses as a compositional object the Schoenberg voice memo that we encountered previously in *Voices and Piano*, but in this case taking a markedly different approach to setting the speech recording. Instead of adding a musical pitch and temporal framework to a recording of speech, Ablinger attempts to fit a musical instrument—the piano—into the pitch and temporal framework of speech without using the original voice in the musical work. Ablinger places the information from the FFT analysis of the recorded speech on what he calls a "grid" of time and pitch, and then, using a program designed by Thomas Musil, converts the acoustic information into midi-data in semitones. An interpretation of the recorded voice is then performed by a computer-controlled player piano crafted specifically for Ablinger by Winfried Ritsch.

Ablinger describes his work in *Quadraturen* as "phonorealism," the aural equivalent of the visual photorealism, wherein an artist reproduces a photograph as realistically as possible in another medium such as drawing or painting.[40] The result is a work in which the featured instrument attempts to closely replicate speech. As with photorealist artworks, phonorealism intentionally does not quite capture the original source but creates a sort of fantastical approximation.

[40] Peter Ablinger, "Peter Ablinger - Quadraturen," July 28, 2016, http://ablinger.mur.at/docu11.html.

How does Ablinger use a musical instrument to sketch the acoustic nuances of speech in a way that we've not seen in previous examples? Here the piano must embody the voice, not accompany it (as in "Çé ça qu'on va faire!") and not approximate it through traditional musical frameworks (as in "Bonnie Barnett" or Tenney's "Song"). In terms of a temporal framework, this means that note attacks come much faster in order to convey the rapidly changing pitch content of the consonant attacks and vowel formants. For *Quadraturen III*, the grid is often so fast that while attacks are still perceptible, they come too quickly for the listener to arrange and identify as a tactus. In "A Letter from Schoenberg," note attacks sound around every 50 milliseconds, much faster than boundary of 100 milliseconds that Justin London identifies as "the shortest interval that we can hear or perform as an element of [a] rhythmic figure."[41]

Turning to the frequency axis of Ablinger's "grid," which is ordered by these quick attacks, each note on the piano is a discrete pitch. Instead of sliding up or down in frequency as in regular human speech, a new attack must be made to produce a new higher or lower pitch—a constraint of the particular instrument that Ablinger has chosen. Pitches are limited to the ones afforded by twelve-tone equal temperament.

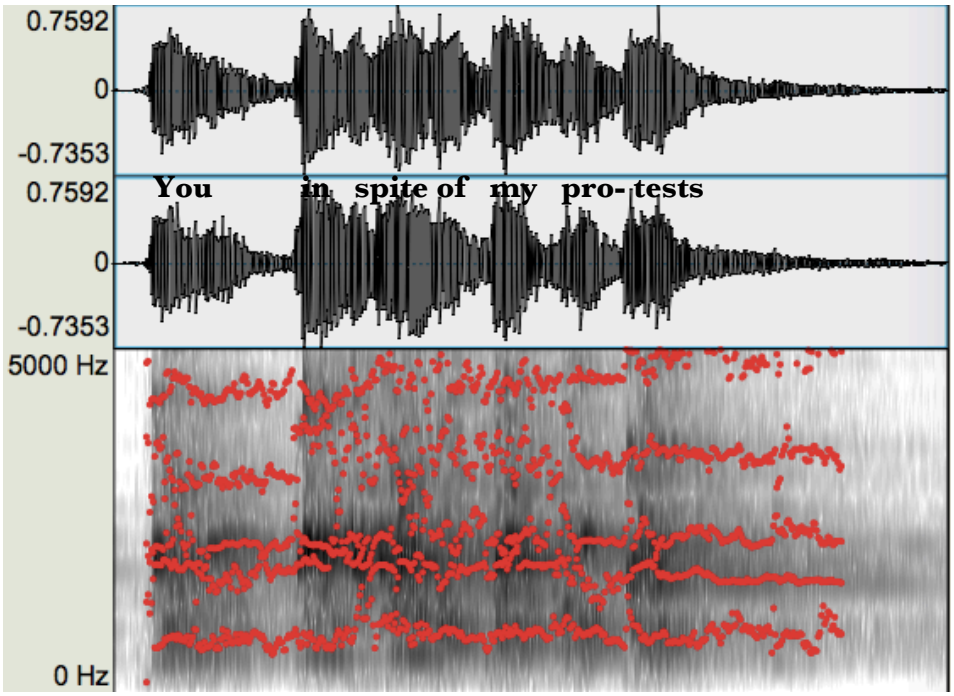
What results aurally in the piece is a perceptual blurring of many quickly-sounded discrete pitches into a larger continuous unit—a syllable. Here a waveform and spectrographic analysis can help illustrate this blurring by visualizing information about timing, volume, and pitch (see Figure 3.13). Figure 3.13a shows such an analysis of the phrase «you in spite of my protests» as played by the piano in "A Letter from Schoenberg." The waveform is included in order to more easily display onsets of syllables. The spectrogram shows the frequency content with formants marked in red. For comparison, Figure 3.13b shows an analysis of Schoenberg's original voice memo. The darker

[41] London, *Hearing in Time*, 27.

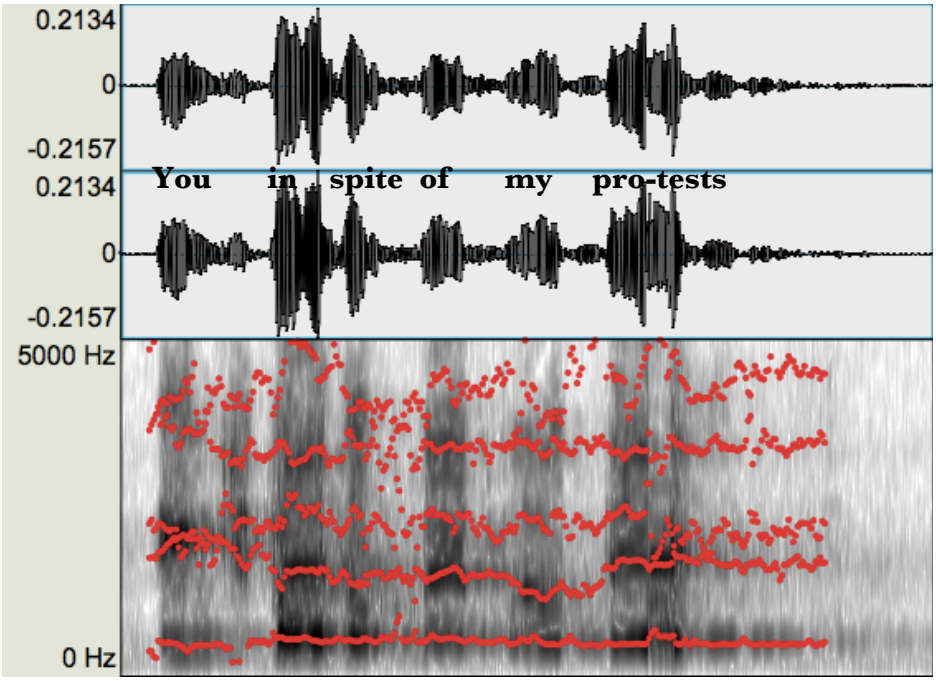
vertical bands show points of higher energy intensity, vowels, while the lighter bands show the vocal tract restriction of consonants. While these contrasts are clearer in 3.13b, the onsets of vowels are nonetheless certainly visible in 3.13a (especially at the onset of «in» and «-test»). Aside from the second formant of «you», which is noticeably lower in 3.13a than in 3.13b, the formant markings of each example sound in about the same range of frequencies. The formants in 3.13a, however, tend to form somewhat more jagged lines than 3.13b. This is due to the changes of discrete pitch, however quick, in the piano instead of the continuous pitch alterations of the voice. The relative phonetic legibility of the "piano speech" is remarkable, however, given the acoustic spectrum accompanying each pitch struck by the keys. In addition to the noise of the hammer striking the string, harmonics sound above the fundamental. While the fundamental frequencies of each note are standing in for the formants of speech, additional frequencies are resonating from the piano, resulting in a much more complex set of frequencies than in the voice itself. Figure 3.13 demonstrates that while the piano-speech from "A Letter from Schoenberg" is indeed some sort of enhanced, non-human version of speech, it nonetheless carries much of the acoustic information of the human speech including timings, volume changes, and vowel formants.

Figure 3.13. Wave form and spectrographic analysis, «you in spite of my protest»

a. "A Letter from Schoenberg," *Quadraturen III*, 0:00–0:05



b. "VR48: Brief, Los Angeles, An Ross Russell, New York," Arnold Schoenberg Center, 0:00–0:05



Although the original recorded voice is not featured in the piece, the influence of the speaker on the composition is undeniable. In "A Letter from Schoenberg," we may find the trace of the human voice, but with major parts of its humanity—including body, vocal timbre, its temporal simultaneity with movements in the body that produce it[42]—stripped away, leaving just some semblance of its frequency and time-durational parameters intact. This frequency and time-durational information derived from the voice reflect the limitations of the piano: the overcoding of the frequency information brought on by the overtones of the instrument and the undercoding of the temporal information brought on by need to attack discrete pitches. Not all of the voice's humanity is lost, however. The piano-synthesized speech is still recognizable as some sort of speech utterance, and its semantic content may even be understandable if heard while reading the text of Schoenberg's voice memo, which Ablinger provides.[43] "A Letter from Schoenberg" is meant to blur the line between human (the voice) and machine (the player piano) by producing sounds that are recognizable as speech (if not intelligible semantically) without a human vocal presence.

The musical works discussed in this section all bring to the fore the differences in organizational structure between speech and music. "Bonnie Barnett" and "Çe ça qu'on va faire!" position speech as dissonant against musical frameworks. In doing so, they confront the

---

[42] N. Katherine Hayles, "Voices out of Bodies, Bodies out of Voices: Audiotape and the Production of Subjectivity," in *Sound States: Innovative Poetics and Acoustical Technologies*, ed. Adalaide Morris (Chapel Hill: University of North Carolina Press, 1997), 76 envisions temporal simultaneity of phone calls and radio as creating a "phenomenology of presence" that allows listeners to feel connected to mediated, disembodied voices.

[43] Ablinger refers to "A Letter from Schoenberg" as a "reading piece with player piano" (Peter Ablinger and Aljoscha Hofmann, "A Letter from Schoenberg," Peter Ablinger, September 31, 2002, http://ablinger.mur.at/txt_qu3schoenberg.html). Again, this brings up the comparison with sine-wave speech, another situation in which listeners are able to decipher linguistic content from a non-human sound source. Robert E. Remez et al., "Speech Perception without Traditional Speech Cues," *Science* 212, no. 4497 (1981): 949 found that sine-wave speech is more readily understood by listeners when they are given linguistic source material.

listener with distinctly two distinctly different modes for communicating ideas. In "Bonnie Barnett," this comparison is made directly with the notion of smoothness in both speaking voice and music. "Çe ça qu'on va faire!" plays up the lack of a stable scalar organizational system in speech by positioning the clarinet-doubled monologue against a tonal, metrically regular groove. "A Letter from Schoenberg" seeks to distance the piano from traditional musical frameworks as much as possible in order to replicate the sounds of speech. But by virtue of its instrumentation, it's impossible to completely dissociate these sounds from music even if they require a different kind of listening than do traditional musical gestures. Going forward into Chapters 4 and 5, I will more deeply investigate the different ways communicating ideas that speech and music offer the listener and examine how these systems of communication work together or clash.

Speech as a system of musical organization

In the music discussed in this chapter, speech still informs the shape of musical gestures, but we see a higher propensity for irregularity or lack of robustness in musical temporal or pitch frameworks than we heard in the music of Chapter 2, which allowed for the alteration of speech in some way to suit the composer's chosen musical frameworks.

In terms of rhythm and meter, in *Pop Titles "You"* we heard a sort of minimal tactus-level framework that expanded to a hierarchically arranged metrical framework in conjunction with a repeated linguistic phrase structure. We can parse "Ants in My Eyes Johnson w/Drums" into recognizable formal sections based on each new drum pattern, but the main gesture toward coherence is in the repetition of a single spoken phrase and its corresponding musical motive at the beginning and middle point of the track. In the "Arnold Schoenberg" movement of *Voices and Piano*, it takes a while for the meter to become salient for the listener.

In terms of pitch, the only works discussed in this chapter that have a clear tonic are "Mrs. Morris" and "Çe ça qu'on va faire!" (although the latter flirts with atonality in the melodic layer). "Arnold Schoenberg" from *Voices and Piano* tends toward free atonality. "Song" from *Song 'n' Dance for Harry Partch* uses a 72-tone microtonal system, the fine-grained nuances of which only the most experienced microtonal listener could hear and grasp. Finally, despite using the resources of 12-tone equal temperament, "A Letter from Schoenberg" from *Quadraturen III* is a sort of speech synthesis by instrumental means more so than a musical work.

While the musical examples in this chapter have different relationships to musical pitch and rhythmic organization, which are a result of their composers' individual compositional approaches, these examples nonetheless illustrate ways in which composers have used a handful of strategies to musicalize speech. In each, the result draws the listener's attention to musical pitch and rhythm, often pushing the listener to hear these features in the speech itself, pushing the listener toward a musical interpretation of the sonic features of speech.

By now you may have noted that the musical pieces I've analyzed come from a variety of disparate musical idioms. You may have noted that I'm trying to theorize and make connections within a corpus of music that does not have clear stylistic expectations on which to build such a theory.[44] A musician drumming along to a cartoon in a YouTube video seemingly has little to do with the world of contemporary art music composition. So what *do* these pieces have in common?

---

[44] Lerdahl and Jackendoff assert that the goal of music theory is to be a "formal description of the musical intuitions of a listener who is experienced in a musical idiom." By emphasizing the musical idiom, Lerdahl and Jackendoff can focus on identifying listener expectations based on musical intuition: "knowledge that enables [the listener] to organize and make coherent the surface patterns of pitch, attack, duration, intensity, timbre, and so forth." And yet I'm not sure it's possible to develop a uniform set of expectations for the music I've covered so far. This dissertation is concerned instead with analysis as a means of elucidating compositional strategies that work across varying musical idioms.

On a surface level, they each use recordings of speech, and they was created within a timeframe that might reasonably be grouped as one historical unit, a sort of long late-20th century that is informed by technological resources. Furthermore, I think what ties these pieces together is an aesthetic stance rather than any sort of musical idiom. Certainly, though, we can make connections in compositional strategies across these disparate stylistic examples, as I have done in this and the preceding chapter.

At the end of the preceding chapter, I proposed that the musicalization of speech is born out of the aesthetic change in music in the twentieth century that drove composers to broaden the range of sounds and actions that are considered musical. I wish to build on that idea here and suggest that for some composers the recorded speaking voice functions as a new, novel system of musical organization, one that they could apply through a variety of stylistic lenses. This is a sort of middle-ground approach to the creation of new music: composers may well draw on existing stylistic idioms (minimalism, rock, jazz, etc.) but they may also draw on the acoustic information of speech to compose music that is innovative within these idioms.

Historical groundwork for speech as a compositional object

Before I move onto a more focused analysis of the power relationships at work in musical pieces based on recorded speech, I'd like to further explore the historical transmission lines that inform this treatment of the voice in music. I see a handful of musical movements and experiments that laid the groundwork for using speech as a compositional object. All of these milieus notably use technology as a means for metaphorically "cracking open" the voice, utilizing it in new ways within musical contexts.

Music of the past century or so has been characterized by experiments in sonic organization—atonality, serialism, found objects, *musique concrète*, and chance music to name but a

few. The use of recorded speech as a compositional object could be interpreted as an extension of the chance music aesthetic: the pitch and rhythm of the compositional object are beyond the composer's control, and the composer must make choices about the structure of the music based on the sonic information she is given. The technologically mediated found objects of John Cage's *Imaginary Landscapes No. 4* (1951), for instance, could be seen to be a predecessor to the works found in this dissertation. In this piece, Cage has performers control 12 radios bouncing between stations in a sort of polyphony. The performers follow notation the governs the temporal duration for each "instrument" (radio), but whatever sounds happen to be coming from the radios at the moment of performance become the sounds of the piece, leaving the aural result the chance. Like works for recorded speech, *Imaginary Landscapes No. 4* asks listeners to hear new musical relationships between found sounds, some of which are recorded. And it is worth noting that some of the composers discussed in this dissertation—specifically, James Tenney and Steve Reich—have a direct connection to John Cage.

Research into speech perception and synthesis has likewise informed the musical approach to the recorded speaking voice, questioning how we hear speech and opening the possibility for artistic experimentation with speech by demonstrating that humans are to some extent capable of hearing and identifying speech-like qualities in sounds that aren't human speech. Musical works (especially on the more institutional, contemporary art music end of the spectrum) that use recorded speech as a compositional object are a sort of natural extension of music-linguistic experimentation begun by avant garde composers in post-WWII Europe. In her book *Electronic Inspirations*, Jennifer Iverson links the composition of works such as Mauricio Kagel's *Anagrama* (1957/58), Luciano Berio's *Thema (Omaggio a Joyce)* (1958), and John Cage's *Aria* (1958) to innovations in military technologies and discourses in the Cold War era, particularly

phonetics and speech-synthesis research.[45] Iverson speculates that what attracted like composers Kagel, Berio, and Cage to experiment with phonetics was that the composers believed that "phonetics and speech-synthesis insights could lead to a music that was *truly* new."[46]

Other threads of inquiry have likewise yielded results that inform the musicalization of speech sounds. Research at Bell Labs in the 1950s resulted in early attempts at singing speech synthesis systems, yielding a successful collaboration between composer Max Mathews and researchers John L. Kelly and Carol C. Lochbaum in the early 1960s.[47] In 1981, Robert E. Remez and his colleagues determined that sine-wave speech, in which three sinusoidal tones replicate the formant frequencies of a natural speech utterance, is understandable as speech when listeners are given the sentence that is being replicated but may not even be processed as speech when listeners are unaware of what the sound is meant to be.[48]

Composers and musicians are attracted to the use of speech recordings as a compositional object because it allows for innovative approaches to composition even when using musical frameworks from existing styles. Such an attitude is highly conditioned by technological developments in speech synthesis. Technology allowed composers and musicians to

---

[45] Jennifer Iverson, *Electronic Inspirations: Technologies of the Cold War Musical Avant-Garde* (New York, NY: Oxford University Press, 2019), 167–93. Iverson traces these innovations back to Homer Dudley's vocoder, first produced in 1935, which took human speech, analyzed its electronic signal into a set of frequency bands, and then synthesized it by passing only the information necessary for intelligibility through two sources of noise (a buzz and a hiss) to create a synthetic reproduction of the voice. For more on the vocoder, see also Homer Dudley, "The Automatic Synthesis of Speech," *Proceedings of the National Academy of Sciences of the United States of America* 25, no. 7 (1939): 378; Mara Mills, "Media and Prosthesis: The Vocoder, the Artificial Larynx, and the History of Signal Processing," *Qui Parle* 21, no. 1 (2012): 107–49.

[46] Ibid., 181.

[47] See Perry R. Cook, "Singing Voice Synthesis: History, Current Work, and Future Directions," *Computer Music Journal* 20, no. 3 (1996): 38.

[48] Remez et al., "Speech Perception without Traditional Speech Cues," 947–50. See also Brian Roberts, Robert J. Summers, and Peter J. Bailey, "The Perceptual Organization of Sine-Wave Speech under Competitive Conditions," *The Journal of the Acoustical Society of America* 128, no. 2 (August 1, 2010): 804–17

recontextualize the voice, to hear it differently, and to unseat assumptions about what both speech and music entailed.

I see another precursor in the latter half of the 20th century with experimentation with using non-musical sounds as compositional objects. James O'Callaghan refers to this approach as "mimetic instrumental resynthesis," a type of composition in which sounds from field recordings are transcribed for acoustic instruments. O'Callaghan traces this technique first to François-Bernard Mâche's musical works of the 1960s and goes on to discuss this idea in relation to the spectral music movement, which began in the 1970s at IRCAM and has branched out over the past half-century.[49] Spectral and post-spectral music's use of the voice as a compositional object is another way of recontextualizing and "cracking open" the voice. Spectral music composers have long been interested in naturalism, using sounds found in the natural world as the basis for musical experiments in the line between timbre and harmony.[50] This interest in the natural is, however, tempered by a highly technological compositional practice aided by advances in technology that made possible the analysis and orchestration of complex sounds. Spectral music is deeply rooted in computer and electronic technologies similar to those that enabled spectral analysis and a wide range of compositional processing techniques. These include instrumental synthesis, computer-aided orchestration, and live processing, all of which have helped composers investigate the spectral content of complex sounds and create new, technologically-mediated musical interpretations of them.[51] Spectral composers have used these technologies to experiment

[49] See also James O'Callaghan, "Mimetic Instrumental Resynthesis," *Organised Sound* 20, no. 2 (August 2015): 231–40.
[50] See Idem., "Spectral Music and the Appeal to Nature," *Twentieth-Century Music* 15, no. 1 (February 2018): 57–73.
[51] For spectral composers' reflections on working with complex sound objects, see Tristan Murail, "The Revolution of Complex Sounds," *Contemporary Music Review* 24, no. 2–3 (April 2005): 121–35 and Gérard Grisey and Joshua Fineberg, "Did You Say Spectral?," *Contemporary Music Review* 19, no. 3 (January 2000): 1–3.

with recomposing natural sounds from the acoustic world, e.g., the forceful E2 of the trombone in Gérard Grisey's *Partiels* (1975), the cathedral bells of Jonathan Harvey's *Mortuos Plango, Vivos Voco* (1980), the piano tones of Tristan Murail's *Désintegrations* (1982), and so on.[52] In the works by Tenney and Ablinger discussed in this chapter, we see the composers relying on acoustically-grounded spectral analyses to produce detailed information about the voice and using these analyses to make compositional choices about pitch and temporal content in their music, an approach that clearly has roots in the spectral approach to musicalizing complex sounds.

In the world of popular music, a blurring of the lines between speech and song has allowed for experimentation with the recorded voice. This experimentation that relies less explicitly on technology than the previous examples, although the importance of the *recorded* nature of this music should not be overlooked. Popular music has opened up new avenues for aestheticizing the voice, reconfiguring what song is. Steven Rings has explored a range of examples from popular music in which the singer speaks in the song, resulting in speech being positioned as "a marked figure against the unmarked ground of more conventionally coded musical sound."[53] In these examples, music serves to aestheticize the speaking voice. Rap is another such nexus of speech and music, employing the hierarchically organized rhythms of

---

[52] For deeper analyses of how these works are constructed from their complex-sound compositional objects, see Francois Rose, "Introduction to the Pitch Organization of French Spectral Music," *Perspectives of New Music* 34, no. 2 (Summer 1996): 6–39; Robert Hasegawa, "Gérard Grisey and the 'Nature' of Harmony," *Music Analysis* 28, no. 2–3 (2009): 349–371; Jonathan Harvey, "*Mortuos Plago, Vivos Voco*: A Realization at IRCAM," *Computer Music Journal* 5, no. 4 (1981): 22–24; Michael Clarke, "Jonathan Harvey's *Mortuos Plango, Vivos Voco*," in *Analytical Methods of Electroacoustic Music*, ed. Mary Simoni (New York: Routledge, 2006), 111–143; Rozalie Hirz, "Frequency-Based Compositional Techniques in the Music of Tristan Murail," in *Contemporary Compositional Techniques and OpenMusic* (Paris: IRCAM, 2009), 104–18.

[53] Steven Rings, "Speech and/in Song," in *The Voice as Something More: Essays toward Materiality*, ed. Martha Feldman and Judith T. Zeitlin (Chicago: University of Chicago Press, 2019), 39.

music while relying less on fixed pitch.[54] Serge Lacasse argues that popular singing voice itself draws inspiration from everyday speech, examining how paralinguistic features of popular singing like whispers and breathing effects enrich musical expression.[55] On the flipside of this, pitch-correcting technologies like Auto-Tune have paved the way for play between speech and music by providing a digital tuning mechanism for the voice, a musical instrument that Catherine Provenzano casts as notorious for being unreliable at keeping in tune in the same way as instruments that can produce neatly fixed pitches, resulting in a kind of forced standardization when subjected to Auto-Tune.[56] In all of these examples, traditional notions of a singing voice distinct from the speaking voice start to break down.

This constellation of milieus and technological interventions—some aware of each other and some independent—have influenced in various ways in which speech has been used as a compositional object. While music and words have been put together throughout human history in the form of poetry and song, Western musicians—working in both popular and (post)modernist traditions—since the mid-20th century have sought to enhance and reconfigure this relationship between music and words through innovative means. Musical works that use speech as a compositional object are an attempt at utilizing technology in order to set the human

---

[54] For an analysis of the pitch features of rap along the lines of speech, see Noriko Manabe, "We Gon' Be Alright? The Ambiguities of Kendrick Lamar's Protest Anthem," *Music Theory Online* 25, no. 1 (May 2019), http://mtosmt.org/issues/mto.19.25.1/mto.19.25.1.manabe.html: [7.1]–[7.11], which examines the intonation of Fabolous and Pharrell Williams in Kendrick Lamar's "We Gon' Be Alright," positing how differences in the two rappers' intonational deliveries might affect the listener's interpretation of the meaning of their statements.

[55] Serge Lacasse, "The Phonographic Voice: Paralinguistic Features and Phonographic Staging in Popular Music Singing," in *Recorded Music: Performance, Culture and Technology*, ed. Amanda Bayley (Cambridge, UK: Cambridge University Press, 2010), 225–251.

[56] Catherine Provenzano, "Auto-Tune, Labor, and the Pop-Music Voice," in *The Relentless Pursuit of Tone: Timbre in Popular Music*, ed. Robert Wallace Fink, Melinda Latour, and Zachary Wallmark (New York, NY: Oxford University Press, 2018), 163.

voice to music in new and novel ways, an act of experimentation brought on by the affordances of modern recording technology.

Part II: Musical Works for Recorded Speech as Collaborative Acts

Lawrence Kramer argues that the presence of a voice in music (even without a text) "activates a set of human relationships that an instrumental performance can only signify."[1] The singing voice conjures in the listener the idea that that voice is a corporeal medium, creating a space of potential or virtual intersubjectivity between the bodies of the vocalist and the listener.[2]

Along these lines, in this second part of the dissertation I position the speaker and the composer as agents in the creation of musical works for recorded speech, imagining these works as an attempt by composers at a sort of collaborative effort, albeit an uneven one dominated by the composer. The speaker contributes emotional expression, linguistic meaning, and a sense of embodied sound, while the composer exerts ultimate authorial control and builds on the sonic resources provided by the speaker. I propose that the use by the composer of others' speaking voices is often born out of a desire by the composer to relinquish some compositional control. Nonetheless, these works represent only a gesture toward collaboration and not true collaboration because they do not support a back-and-forth or an equal distribution of labor between the speaker and the composer.

Audio recording's ability to capture human voices poses something of a problem for the speaker's subjectivity. Recording technology allows voices to be preserved, repeated, and disembodied, to transform the interiority of the voice in the body into an external cultural function that can be preserved and used by future generations.[3] Simply put, audio recording reconfigures the bounds of the self.

---

[1] Lawrence Kramer, *Musical Meaning: Toward a Critical History* (Berkeley: University of California Press, 2002), 54.
[2] Ibid.
[3] See Jonathan Sterne, *The Audible Past: Cultural Origins of Sound Reproduction* (Durham: Duke University Press, 2003), 287–333.

In the music discussed in this dissertation, the voice originates from a natural source—a human being, but as it is recorded and played back it is mediated by mechanical or digital means. Many such pieces go beyond simple playback of the recorded voice, using digital processing, splicing, or even instrumental representations of the voice's spectral information to alter vocal delivery. In addition, the sound of the recording itself factors into some speech recordings. In Jacob TV's "Billie" (discussed further in the following chapter) we can hear tape hiss from the recording. In "Fidelito/La Revolución y las Mujeres" from Peter Ablinger's *Quadraturen III*, the entire recording—including record pops and a screaming crowd—is transcribed for the player piano. These sounds call attention to the recorded medium: there is not just a speaker and music, the simple fact of recording the voice introduces a level of mediation between the speaker and the music, and with it the potential to disembody the voice. As a result, the subject of the speaking voice is highly reconfigured—mechanically or digitally mediated but still rooted in some natural reality. Here, sonic virtuality—which results from any artificial configuration of sound (sounds of fictional entities, sounds with no identifiable source, hyperreal sounds that represent real-life sounds with artifactual ones, or real sounds that are propagated by artificial equipment)—complicates traditional notions of subjecthood and agency.[4]

I see works for recorded speech as attempts at virtual collaborations between the speaker and the composer along the lines of the expanded notion of agency envisioned by Jason Stanyek and Benjamin Piekut in their 2010 essay on posthumous musical collaborations. They argue that agency should not be relegated to one person in the present/future, but rather that agency can flow in multiple temporal directions and that agency is created by a network of actors including

[4] Tom A. Garner and Mark Grimshaw, "Sonic Virtuality: Understanding Audio in a Virtual World," in *The Oxford Handbook of Virtuality*, ed. Mark Grimshaw (New York: Oxford University Press, 2014), 367–68.

the living, the dead, machines, and discourses.[5] For them, agency is delineated by an agent "having an effect" in a collaborative effort, allowing for the inclusion of humans that are present only in voice.[6] For Stanyek and Piekut, "personhood is not equivalent to a lone body, but is distributed among and articulated with other entities that are textual, technological,…and affective."[7] Stanyek and Piekut's model of agency and personhood allows us to consider the speaker as more than a compositional object, to take into account how the speaker's individual voice and body affect the musical work and its composer.

Ultimately, I wish to frame this collaboration as something of a power play between the voices (literal and metaphorical) of the speaker and the composer. The composer's voice traps the speaker's voice into musical frameworks, adding new layers of meaning. The speaker's voice provides a fixed sound object. Its speech utterances are organized along different principles than those through which the composer organizes her materials, and rely on different expressive resources that exert force on those of music. The following chapter will examine how the expressive potential of recorded speech empowers the speaker in this collaborative act, focusing particularly on how expressive speech adds meaning and how this meaning complemented by musical expressivity. Chapter 5 then focuses on the composer's role in crafting a musical structure that suits the sonic features of a speech recording. I am particularly interested in what each of these agents bring to the table in terms of communicating meaning, and argue that the resulting works are a melding of linguistic, expressive, and musical meaning.

---

[5] Jason Stanyek and Benjamin Piekut, "Deadness: Technologies of the Intermundane," *TDR (1988-)* 54, no. 1 (2010), reprinted in *The Sound Studies Reader*, ed. Jonathan Sterne (New York: Routledge, 2012), 304–323.
[6] Ibid., 307.
[7] Ibid., 308.

Chapter 4: Expressivity in Speech and Music

"Whatever music may 'mean,' it is in no sense comparable to linguistic meaning."[1] Fred Lerdahl and Ray Jackendoff make this provocative claim in the introduction to their 1983 *A Generative Theory of Tonal Music*, a study which seeks to craft a system of musical grammar cast on a linguistic model. They elaborate,

> there are no musical phenomena comparable to sense and reference in language, or to such semantic judgments as synonymy, analyticity, and entailment. Likewise there are no substantive parallels between elements of musical structure and such syntactic categories as noun, verb, adjective, preposition, noun phrase, and verb phrase. Finally, one should not be misled by the fact that both music and language deal with sound structure. There are no musical counterparts of such phonological parameters as voicing, nasality, tongue height, and lip rounding.[2]

It is this final point about sound structure that interests me most. Here Lerdahl and Jackendoff begin to consider paralinguistic, or nonverbal, vocal components of speech. Because I am interested in intersections and conflicts between music and natural language as means of communication in what follows, I want to explore the expressive power of such paralinguistic features in the context of musical works that use recorded speech as a compositional object. What can music do when it is juxtaposed with spoken natural language and its particular mode of expressivity? How can music reflect these qualities in its own system of communication? What is gained or lost by translating the expressive features of speech into musical gestures? What's the role of the individual speaker in the communication of emotions or moods? These are the questions that motivate the analyses undertaken in this chapter.

First, a few clarifications. What do I mean by expressivity? Expressivity and the related act of expression are slippery terms, but on a basic level they relate to how music communicates

---

[1] Fred Lerdahl and Ray S. Jackendoff, *A Generative Theory of Tonal Music* (Cambridge, Mass.: MIT Press, 1983), 5.
[2] Ibid., 5–6.

beyond simply projecting sounds to the listener: they are sonic gestures that suggest emotions or ideas to the listener.

Where then do we locate expressivity in speech or musical utterances? The answer is complicated. Expression is a feature of both speech and music that is important in communicating moods and emotions, but is often pushed aside for more easily studied parameters. Locating which parameters of speech or music convey emotions is nonetheless a difficult task. Although studies have found the listeners are quite good at correctly recognizing emotions like joy, anger, sadness, fear, and tenderness in speech, researchers have struggled to pinpoint the exact acoustic patterns for these emotions.[3] In music, scholars have proposed that the interpretation of emotional content music is related not only to features of the sound but also to expectancy, arbitrary association, and iconic signification.[4] Both music and speech prosody—vocal expression of speech, particularly with regards to intonation and rhythm—are, according to Gabriela Ilie and William Forde Thompson, "acoustic stimuli that convey emotional meaning through variation in pitch height, pitch contour, intensity, rate (tempo), timbre, and rhythmic grouping."[5] Stephen Davies gives a slightly different formulation when he argues that the semantic content of utterances in both speech and music can be affected by "pitch, rhythm, tempo, accent, phrasing, attack, and decay."[6] From these examples, we see that scholars have a somewhat slippery grasp of which sonic features affect expressivity.

---

[3] Ingrid Verdyckt and Dominique Morsomme, "Perception of Emotion in Impaired Facial and Vocal Expression," in *Emotions in the Human Voice*, ed. Hans von Leden and Luigi Anolli (San Diego: Plural Publishing, 2008), 134.
[4] Patrik N. Juslin and Petri Laukka, "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?," *Psychological Bulletin* 129, no. 5 (2003): 774.
[5] Gabriella Ilie and William Forde Thompson, "A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect," 319.
[6] Stephen Davies, *Musical Meaning and Expression* (Ithaca: Cornell University Press, 1994), 2.

Linguists Mark Tatham and Katherine Morton describe expression in speech as "a manner of speaking, a way of externalizing feelings, attitudes, and moods—conveying information about our emotional state." Expressive speech "gives us information, other than the plain message, about the speaker, and triggers a response in the listener."[7] A crucial facet of speech expression is intonation, the variation in spoken pitch. Linguistic intonation conveys syntactic, pragmatic, and emphatic information and also signals prosodic grouping patterns.[8] Going forward I will be particularly interested in intonation's effect on pragmatic information, the use speakers make of words relative to context.[9]

Although the preceding literature review suggests that expression is a somewhat slippery, hard-to-define phenomenon, we can nonetheless see potential points of convergence between expressive speech and musical expressivity. Nineteenth-century music theorist and scientist Hermann von Helmholtz suggested that vocal expression may have influenced the development of musical expression, claiming that "an endeavor to imitate the involuntary modulations of the voice, and make its recitation richer and more expressive, may therefore possibly have led our ancestors to the discovery of the first means of musical expression."[10] More recently, Serge Lacasse has argued that popular singing idioms are derived from everyday speech, focusing on how paralinguistic features of speech like whispers, creaky voice, and falsetto are manifested in

---

[7] Mark Tatham and Katherine Morton, *Expression in Speech* (New York: Oxford University Press, 2004), 39.
[8] Aniruddh D. Patel, *Music, Language, and the Brain* (Oxford: Oxford University Press, 2008), 186. I'm setting aside the complications raised by languages that differentiate words through pitch inflection (as happens in a number of Asian and African languages) because the languages used in the speech recordings discussed in this dissertation are not tonal languages.
[9] Jeffrey C. King and Jason Stanley, "Semantics, Pragmatics, and the Role of Semantic Content," in *Semantics versus Pragmatics*, ed. Zoltan Gendler Szabo (Oxford: Oxford University Press, 2005), 111–161.
[10] Hermann von Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music.*, trans. Alexander John Ellis (New York: Dover Publications, [1863] 1954), 371.

popular music.[11] Greg Urban, in his study of ritual wailing as an expression of sorrow, identifies four icons of crying in ritual wailing: cry breaks that phonologically form sob sounds, voiced inhalation, creaky voice, and falsetto vowels.[12] On a very local level, limited to one form of emotional expression, Urban explains how the spoken vocal expression has formed the basis of the musical one. Patrik Juslin and Petri Laukka, for their part, tie the vocal expression of emotion to its evolutionary origins and suggest that the origins of music's emotional expressivity can be traced back to its use in human social activities.[13]

The study of expressive speech (as well as expression in music!) reminds us that there's more to communication than simply understanding the meaning on the most basic level. This can happen within the communicative medium itself, or as a confluence of multiple communicative media. In speech, beyond the words, intonational or paralinguistic features add to our pragmatic interpretation and understanding of an utterance. Speech can also be paired with other means of communication in order to enhance meaning. In comparing speech and gesture as communicative media, David McNeill argues that "When co-expressive speech and a gesture synchronize, we see something that is both simultaneous and sequential…There is a combination of two semiotic frameworks for the same underlying idea, each with its own expressive potential."[14] Similar processes are at work when combining speech and music, each of which come with their own communicative potential. Building on McNeill's work, Lawrence

---

[11] Serge Lacasse, "The Phonographic Voice: Paralinguistic Features and Phonographic Staging in Popular Music Singing," in *Recorded Music: Performance, Culture and Technology*, ed. Amanda Bayley (Cambridge, UK: Cambridge University Press, 2010), 225–251; see 226–7 for a literature review of connections between speech and the popular singing voice.
[12] Greg Urban, "Ritual Wailing in Amerindian Brazil," *American Anthropologist* 90, no. 2 (1988): 389–91.
[13] Juslin and Laukka, "Communication of Emotions in Vocal Expression and Music Performance," 771–75.
[14] David McNeill, *Gesture and Thought* (Chicago: University of Chicago Press, 2005), 91.

Zbikowski proposes that both musical and gestural statements "interact with and shape the story that is told through language."[15] I posit then that affixing music to a recording of speech changes the way we understand the speech, adding additional co-expressive layers of meaning.

Going forward, I will analyze two musical works that use recorded speech as a compositional object as a way to compare how speech and music convey expressive content. These works, "Billie" by Jacob TV and "Perpetual/Pop Titles 'You'" by DJ Spooky, both reconfigure the speaking voice of Black women musicians. This comparison will center the speaker in the musical collaboration, exploring how the expression of the speaking voice interacts with and contributes to the expressivity of the music. In doing so, I wish to interrogate the claim from Lerdahl and Jackendoff that starts off this chapter—that linguistic and musical meaning are two very different things. I argue that these two modes of communicating meaning have the potential, when melded together like this, to work in a complementary way, reinforcing one another and providing a sort of heightened meaning. Paralinguistic components of speech are actually important for linguistic meaning and they connect with the expressive resources exploited by music.

These musical works are built on the qualities of expressive speech of *individual subjects*. The effect just wouldn't be the same if the same phrases were said by a different speaker, or expressed differently by the same speaker. Certainly speakers contribute meaning in their linguistic utterances, meaning that the listener places in context with the music. Moreover, though, they contribute emotions or moods through their expressive delivery of these linguistic utterances. I want to propose that this expressivity is really what draws the listener's attention to

[15] Lawrence M. Zbikowski, "Musical Gesture and Musical Grammar: A Cognitive Approach," in *New Perspectives on Music and Gesture*, ed. Anthony Gritten and Elaine King (Farnham, Surrey: Ashgate, 2011), 81.

the speaker as a subject, as an individual with a pointedly unique voice whose phonological features are underscored and supported by the music. An exploration of expressive speech in music thus brings to the fore the role of the speaker as a human subject in the composition of these works; moreover it brings out how the speaker's individuality is manifested and what force the speaker's voice exerts on the compositional process.

"Billie," Jacob TV

I turn first to "Billie" (2003), composed by Dutch tonal avant-garde composer Jacob TV (b. 1951) for live alto saxophone and a tape "soundtrack" consisting of voice, piano, additional saxophones, bass, and drum machine. This work features the speaking voice of Billie Holiday (1915–1959) excerpted from various interviews over the course of her career. By juxtaposing Holiday's speaking voice from different periods of her life and by adding instrumental backing, TV brings out the virtuosity and musicality of Holiday's speech.

"Billie" is an example of TV's "boombox repertoire," which he calls works for "live instruments with a grooving sound track based on speech melody."[16] TV is particularly fixated on the voices of American pop culture, using samples of the speaking voices from sources such as an advertisement for fitness equipment ("Body of Your Dreams"), a televised sermon by an evangelical preacher ("Resurrection Power"), confrontations from 1990s daytime talk shows ("Heartbreakers"), and interviews by pop culture icons like Marilyn Monroe ("Able to Be") and Chet Baker ("Pitch Black"). Critic Anne Midgette claims that TV's work "engages in a kind of

[16] "Bio," JacobTV.net, accessed September 26, 2019, https://www.jacobtv.net/composer.html.

pop-culture appropriation that has distinct parallels to visual art," likening TV's boombox pieces to Jeff Koons's exploration of kitsch and Tom Wesselmann's use of collage.[17]

Billie Holiday is a particularly apt American pop culture icon for TV's treatment of speech: critics, biographers, and musicologists have often remarked on the musicality of Holiday's speech.[18] Hao Huang and Rachel Huang have taken this comparison one step further: their 2013 essay "She Sang as She Spoke" explores the blurry line between speech and song in Holiday's voice, identifying several aspects of Holiday's singing voice that borrow from speech.[19] More specifically, they point to two phonological features of her singing voice that borrow from expressive speech. First, they identify her use of pitch slides as a speech-like feature that serves as a carrier of emotional content. Second, the soft or "blurred" consonants in her singing voice enhance her diction and emphasize the microtonal inflections of her pitch slides. Huang and Huang end their analysis by comparing three of Holiday's sung performances of "Yesterdays" with a recording of her spoken recitation of the lyrics. They identify Holiday's voice as a site of reciprocity between speech and song, where her manner of speech affects her singing and her style of singing inflects her speech. Jacob TV's "Billie," then, provides another case study to interrogate the mutual influence of speech and song in Holiday's voice.

Turning to the musical work at hand, the vocal content of "Billie" consists of recordings by Holiday from different periods in her life, both speaking and, in the contrasting **B** section only, singing. This gives us an opportunity to consider her *voice* as opposed to just her speech. We can hear a marked change in the quality of her voice over time. Furthermore, the vocal excerpts

---

[17] Anne Midgette, "Dutch Composer Samples Pop Culture and Gives It a Melody," *The New York Times*, May 4, 2007, sec. Music, https://www.nytimes.com/2007/05/04/arts/music/04jaco.html.

[18] See Hao Huang and Rachel Huang, "She Sang as She Spoke: Billie Holiday and Aspects of Speech Intonation and Diction," *Jazz Perspectives* 7, no. 3 (December 2013): 287 for a quick gloss.

[19] Ibid., 287–302.

curated by TV provide quite a range of vocal expression. In considering Holiday's singing, Roland Schleifer points out several instances in which scholars have noted sadness or anguish in Holiday's singing voice, which "flirt[s] with a sense of Holiday's singing as essentially expressive."[20] Holiday's voice, then, in the contrast between speaking and singing provides a site for comparison between how music and speech communicate expressively.

My analysis will focus on several aspects of how musicalizing speech heightens meaning. First I'll consider how the addition of musical frameworks to the sounds of speech prompts different hearings of the pitch contour and rhythm of speech. Next, I'll explore how TV uses the line between speech and song in Holiday's voice as well as the musical accompaniment to highlight the fragility of Holiday's voice late in life. Finally, I'll focus on how TV uses his technological mediation to change Holiday's meaning, both through her expression and her words.

First, a note on the structure of the piece: the form of the piece divides roughly into **ABA'**. See Appendix 4.1 for a diagram that identifies large formal sections, instrumentation, and changes in the textural uses of the voice and saxophone.[21] Sections **A** and **A'** feature Holiday's speaking voice while her singing voice leads the **B** section, particularly in the later portion of this section, which serves as a sort of development of the more reserved musical mood of **B**.

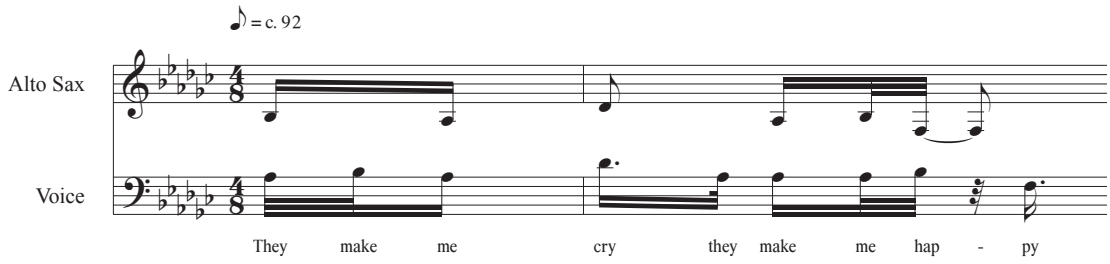
*****

I begin my analysis by focusing on TV's deployment of a single spoken phrase in different musical contexts. As I posited in in my discussion of the intersections of musical frameworks and

[20] Ronald Schleifer, *Modernism and Popular Music: Language and Music in Gershwin, Porter, Waller, and Holiday* (Cambridge: Cambridge University Press, 2011), 160.
[21] Timings are taken from Connie Frigo, "Billie," on *Shining City*, CD (Aalsmeer, The Netherlands: Basta, 2007). Because the piece is performed with tape accompaniment, all performances should have similar timings.

unaltered speech in Chapter 3, the musical setting of a spoken "motive" (in the sense that it is a recurring sonic gesture) can help the listener make connections within and across segments of speech that might not have seem significant until highlighted through musical repetition, and can bring these repeated phrases to structural prominence. In doing so, this repeated spoken phrase becomes a musical motive, which can be analyzed using the same approach I used in Chapter 3. Because of the varied uses of this motive in "Billie," it demonstrates how meaning can be flexible when cast in different musical settings. By choosing to repeat a spoken phrase in different musical contexts, TV creates a link between the semantic and pragmatic content of the excerpt and the musical gestures that accompany it.

Figure 4.1. Sax and voice, score, m. 55, Jacob TV, "Billie," 2:21–2:24



We first hear Holiday say «They make me cry, they make me happy» at 2:21–2:24 followed swiftly by a repetition at 2:29–2:31. Here the saxophone more or less "doubles" the pitches of the speech, reinforcing and regularizing the formant frequencies of the vowels in such a way that it draws attention to the "speech melody" of Holiday's intonation. (See Figure 4.1 for the score.)

We hear this spoken motive again at 4:30–4:38, but in a completely different musical landscape (Figure 4.2). Whereas in the **A** section it was clear that this spoken motive was being sampled out of its original spoken context, here it's part of a story that Holiday tells about one of

her early performing experiences, finally explaining to the listener what makes her cry/makes her happy. (Figure 4.3 shows the text of the spoken portion of the **B** section, 4:03–5:00.) The story is a recollection of the past, and as such the musical accompaniment takes on a more reflective, wistful mood. When compared to the music surrounding the first iteration of this motive in the **A** section, the surface and harmonic rhythms in the **B** section are quite slow, with the bass and synth strings offering harmonic change at most once per measure. The saxophone is no longer tied rhythmically and melodically to the voice. Instead it performs a sort of ad-libbed melody, soaring to a high point as the first iteration of the motive ends in a way that draws attention to this moment.

Figure 4.2. Sax and voice, score, mm. 94–96, Jacob TV, "Billie," 4:30–4:38

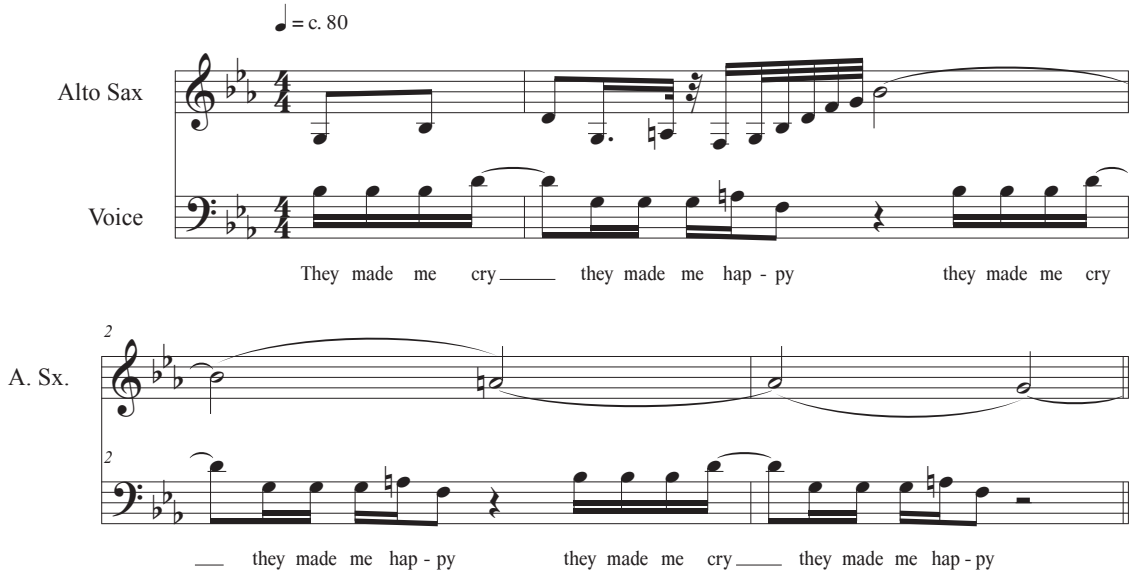


Figure 4.3. Spoken text, Jacob TV, "Billie," 4:03–5:00

I always knew I could sing because I always did sing, but, uh
I always knew I could sing because I always did sing, but, uh
I always knew I could sing because I always did sing, but, uh
I always knew I could sing because I always did sing, but, uh
So I sang and everybody loved me and made about 40 dollars in tips, and I got the job
They made me cry, they made me happy
They made me cry, they made me happy
They made me cry, they made me happy

Because this motive is predicated on a direct repetition of a recording, the phonological content of the motive is the same each time it's heard. There are, however, two notable divergences between these two iterations of the motive in the performance score, which seems to suggest that TV is encouraging the performer and the listener to hear the variation brought to the recorded segment of speech (the "motive") by its shifting musical contexts.

First, whereas in the motive's first iteration the verb is given in its present form («They *make* me cry, they *make* me happy»), in the second iteration the verb is transcribed in the past tense («They *made* me cry, they *made* me happy»). Obviously the switch is predicated on some ambiguity in the recording about which phoneme is actually being produced, but the switch is interpretively fitting—the past tense version of the motive comes in a musical section where the mood is more introspective than in previous sections. In the **B** section the emotional content of the music is thus more clearly in service of the linguistic meaning, setting Holiday's reminiscences to slower, more lyrical music.

Second, TV assigns the motive different rhythms and pitches in the score in the two iterations. The voice is transcribed by TV into notated pitches and rhythms. These are of course a construct—idealized analyses of the sound of speech, transcribed into a musical grid of time and pitch—but they can show us how TV hears or expects the listener to hear the speaking voice in a given musical context. The composer thus seeks to actively shape the musical interpretation

of speech, offering insight into his own hearing of how the fluctuating vowel formant frequencies and aperiodic rhythms of speech fit into his chosen musical frameworks.

Whereas there's a neighbor note B-flat transcribed for the first «make» in the **A** section (Figure 4.1), «they made me» holds steady on A-flat in the **B** section (Figure 4.2). The step up to «hap-» is a whole step to B-flat in the **A** section and a half step to A-natural in the **B** section. Due to differences in tempo, while both recorded excerpts start one beat before a new measure, the leap to «cry» comes on a downbeat in the **A** section and a quarter-beat before the downbeat in the **B** section. These differences remind us that the musicalization of speech is an interpretive act on the part of the composer. Because speech does not typically fall into regular rhythms and discrete pitches, it is up to the composer to determine which of a number of close-but-not-exact musical possibilities will best fit into the chosen musical framework and to find a way of conveying this interpretation to the listener. This is, quite clearly, an appropriation of Holiday's voice and meaning—TV uses his musical setting to manipulate what the listener hears, to enhance Holiday's voice beyond its original utterance. But the fixity of Holiday's voice, the pitch contour of her voice and its delivery across time, mean that the expressive features of Holiday's exert an undeniable inspirational force in shaping the musical possibilities for this phrase.

*****

While the voice-as-recorded is treated by composers as a "compositional object," something to be manipulated at will, for the listener it is still a *voice*, and thus has the potential to carry with it a connection to the speaker. That situation is highlighted in the juxtaposition of speaking and singing in the **B** section of "Billie," a juxtaposition that emphasizes Holiday's ailing body in her later life. Holiday's voice is not only distinctive, but it is also a record of her lived experience, something audible in this late recording. This record of Holiday's then becomes a

part (wittingly or unwittingly) of TV's composition. This highlights how the expressive, pragmatic features of speech are linked to the individual body.

Holiday's singing voice (beginning at 5:00) is taken from "The End of a Love Affair (The Audio Story)," a vocal solo outtake from the recording session for Holiday's penultimate album *Lady in Satin* (1958).[22] Reviewer Derek Ansell, writing about Holiday's delivery on *Lady in Satin*, observes "Although Billie's voice was cracked and strained throughout and you can hear it on every selection, the ravaged voice puts across the pain, sorrow, lost-love feeling, hurt and jazz feel of every note of every song… Billie pours everything (and she didn't have an awful lot left) into interpreting these songs…"[23] In the 1950s, Holiday's health deteriorated due to drug abuse, drinking, and abusive relationships. She died the year after *Lady in Satin*'s release due to complications from cirrhosis.[24] In this recording, we hear Holiday's late style as Tiffany Naiman conceives of it: a form of creativity in which the artist is aware of and contending with loss, sickness, debilitating illness, or a sense of her own mortality.[25]

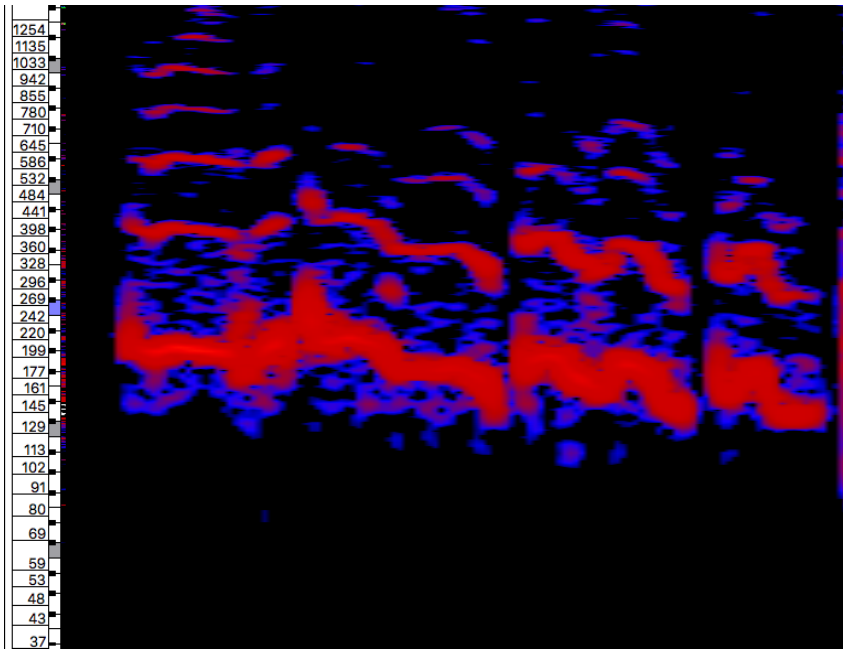
TV sets Holiday's voice in such a way that it plays up this late style. Let me begin by comparing Holiday's late speaking voice with her voice in the decade prior. Figure 4.4 shows a spectrogram of a phrase from an interview that Holiday recorded in 1948.[26] Contrast this with

---

[22] "The End of a Love Affair (The Audio Story)" appears on the reissue Billie Holiday and Ray Ellis and His Orchestra, *Lady in Satin*, CD (New York: Columbia, 1997).
[23] Derek Ansell, "Billie Holiday: Lady In Satin | Jazz Journal," *Jazz Journal* (blog), September 17, 2019, https://jazzjournal.co.uk/2019/09/17/billie-holiday-lady-in-satin-2/.
[24] See Donald Clarke, *Billie Holiday: Wishing on the Moon* (Cambridge, MA: Da Capo Press, 2009) for more on Holiday's personal struggles.
[25] Tiffany Naiman, "Singing at Death's Door: Late Style, Disability, and the Temporality of Illness in Popular Music" (Ph.D., University of California, Los Angeles, 2017), 1–8.
[26] Billie Holiday, Turn Back the Turntable, radio broadcast, July 30, 1948. A recorded can be accessed through *Rare Billie Holiday Interview Pt. 2*, accessed September 28, 2019, https://www.youtube.com/watch?v=bENqqcFDIME.

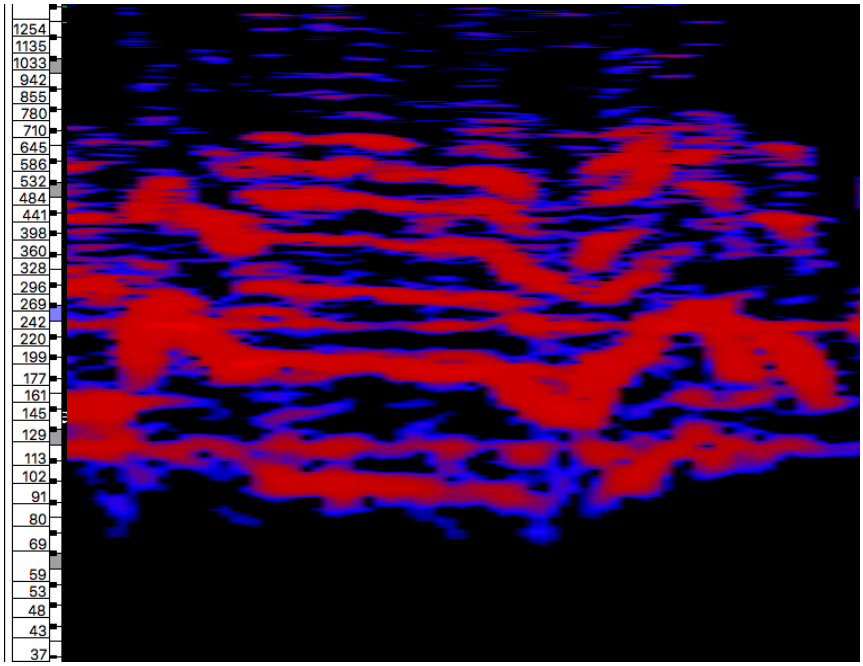
the spoken phrase in Figure 4.5, which is from an interview in 1956.[27] In the 1948 recording, we can clearly see a fundamental and two or three formants above it; this is fairly standard for spectrographic analyses of the speaking voice. In the 1956 recording, however, we see many more frequencies coming through. I attribute this to additional noise, a sort of croaking sound, that has become part of Holiday's voice.

Figure 4.4. «Oh, I believe that was around '39 or '40», Billie Holiday, Turn Back the Turntable (radio broadcast), July 30, 1948



[27] Billie Holiday, Peacock Alley, interview by Tex McCleary, TV broadcast, November 8, 1956. A recording can be accessed through *Billie Holiday Interviewed on Peacock Alley TV Show (1956)*, accessed September 28, 2019, https://www.youtube.com/watch?v=JDl1fkwsDXM.

Figure 4.5. «People didn't know whether I was gonna dance or sing»[28] Billie Holiday, Peacock Alley (TV broadcast), November 8, 1956



TV uses not only Holiday's vocal quality to underscore her late-life personal struggles, but also employs a couple kinds of musical gestures in order to imbue this piece with a sort of melancholia about and hyperawareness of Holiday's deteriorating body. First, the musical setting of Holiday's voice echoes a late-style mood of reflection on the passage of time. Second, the roughness of Holiday's late voice, both singing and speaking, is underscored by several sonic analogs for her damaged voice.

It's particularly notable that with one small possible exception in 2:17–2:21 («And I'd like to do a little tune»), Holiday's singing voice is confined to the development of the **B** section (as set out in Appendix 4.1), which is markedly different in terms of texture, instrumentation, and rhythm than the speech-driven **A** and **A'** sections. **A** and **A'** are driven by a groove that features voice, sax, piano, bass, and drums. The groove is abandoned for the cadenza and first portion of

[28] The last three syllables of this phrase are sampled in "Billie" and transcribed as «Dare to sing».

the **B** section, but the instrumentation remains somewhat thick with voice, sax, synth strings, and bass. When the singing voice enters at 5:00, however, Holiday's singing voice is set in canon with itself with just the saxophone for accompaniment. I hear the canon as a sort of reflective gesture, Holiday immediately looking back on what she's done as her voice shows signs of age and illness. It also draws attention to the voice, giving the listener the opportunity to hear everything twice, thus calling attention to Holiday's bodily struggle to control her voice. Meanwhile, TV's marked reduction in instrumentation signals that this new mode of vocalization is different, special, significant. From 5:00–5:52, the musical mood comes across as sweet and lyrical as only Holiday's singing voice and the saxophone are featured, which could be taken to be a nod to Holiday's enduring friendship and musical partnership with saxophonist Lester Young.

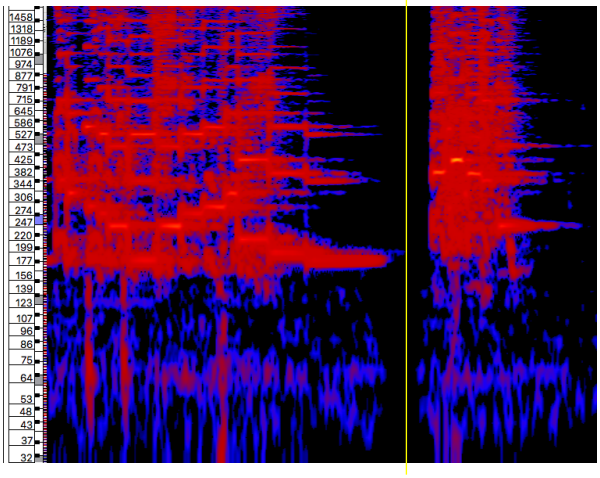
This effect is ruptured at 5:53. We hear a slippage between the singing and speaking voice in 5:53–5:58 as Holiday breaks from singing to shout «I can't hear the band at all». Until now in this section, the listener has been able to imagine the sung voice as a strictly musical performance, but this spoken interjection shatters that illusion, drawing attention to the archival, behind-the-scenes-in-a-recording-session nature of this particular recording of Holiday. We also get the sense that Holiday is not well as she gives a wheezing laugh following this exclamation. At this point, the vocal canon is abandoned, but the saxophone remains to "double" the speaking voice. Notably, the saxophonist employs a growl effect—an extended technique produced by humming into the instrument while playing—as it reinforces the pitches of Holiday's speaking voice.

This saxophone growl serves to enhance the grittiness of Holiday's voice as a sort of musical analog for her vocal quality. As theorized by Lawrence Zbikowski, sonic analogs for dynamic processes in music are on a basic level "the mapping of systematic structural

relationship between disparate domains."[29] Crucially, sonic analogs for dynamic processes correlate a coherent sequence of non-musical phenomena with a musical gesture.[30] Here the deterioration of Holiday's voice is represented in a musical context by a musical effect, which relies on perceivable timbral similarities between the two sounds.

Figure 4.6. Spectrogram of saxophone, Jacob TV, *Billie*, 5:46–5:56

a. dBV^2 scale



b. Linear scale

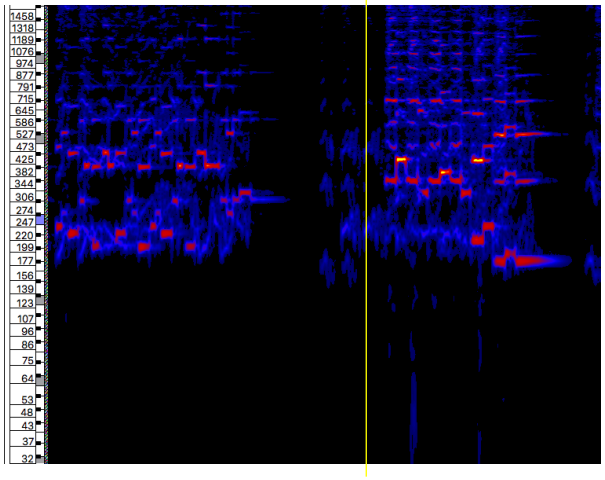


Figure 4.6 shows spectrograms of the growl corresponding to «I can't hear the band at all», preceded by a musical phrase in which the saxophone does not employ a growl effect. The two phrases are separated by a yellow vertical line. I've included two different visualizations, 4.6a using a dBV^2 scale and 4.6b using a linear scale, in order to offer a more complete view of the sonic difference between "ordinary" saxophone timbre and a growl. 4.6a gives us a better sense of the sort of timbral density (the propagation of individual frequencies above a fundamental) while 4.6b makes it much easier to make out pitches. As we can see particularly well in 4.6a, the

[29] Lawrence Michael Zbikowski, *Foundations of Musical Grammar* (New York, NY: Oxford University Press, 2017), 30.
[30] Ibid., 42–47.

growl presents a much thicker, noisier timbre. This comparison shows how the noisier timbre of the sax growl functions as an analog for the noisier timbre of Holiday's failing voice.

Following the sax growl, I find the deterioration of Holiday's voice more aurally apparent. This is due in part to the fact the TV has called attention to her vocal quality, but it also has to do with how TV treats her voice through digital processing. The reverb TV applies plays up the grit of Holiday's voice, while the distortion he introduces creates a timbral sonic analog for the degradation of her voice.

*****

As we move to the **A'** section, TV again uses technology to interact with spoken and musical meaning, this time by splicing and reordering Holiday's speech in a way that strips her utterances of semantic meaning. TV uses technological mediation and control to dismantle Holiday's voice from its original form. He thereby strips it of its ability to communicate in the traditional way we expect speech to communicate, reducing it merely to its pragmatic, expressive elements. This process throws into focus the object-ness of the speech recording and highlights the fundamental inequity of the composer-speaker "collaboration."

This process begins almost imperceptibly in the transition between **B** and **A'**. From 7:34 to 8:04, TV plays with and picks apart a sung phrase, «and the smile on my face», which we've previously heard in the **B** section at 7:17–7:25. The phrase is given first in its entirety from 7:34 to 7:40. It begins with a light distortion added by TV, enhancing the grittiness of Holiday's voice. But when the final word «face» is heard at 7:40, TV drastically changes how the voice is processed, removing the distortion and soaking the sound in reverb. The result introduces a jarring schism between different versions of Holiday's voice: the impression is of two soundworlds that belong to different recordings, different moments in Holiday's life, and different versions of her body. «Face» is then repeated several times, separated from the rest of the phrase, while

unintelligible bits of speech are interspersed. The speech, low in pitch and reordered so that it loses semantic meaning, is a third representation of Holiday's voice in time and space. This section represents a shift from the put-together singing voice of Holiday toward a more fractured version of the self. TV's shifting sound spaces draw attention to the technological mediation of the recorded voice. This is not a pure voice in a single room recorded by a single recording set-up. These are an interpretation of the many voices of Billie Holiday, separated by time, physical space, and the changing quality of recording technology.

In the **A'** section, beginning at 9:14, TV splices Holiday's voice into mere phonemes («to an an», «no s no t»). The voice is accompanied by a minimalist groove in the sax, piano, bass, and drums. The choppy and repetitive musical accompaniment serves to amplify the vacant quality of the speaking voice sans meaning. It bears mention that the reduced materials TV uses here connect with an observation made by scholars, who have pointed to the minimalist aesthetic as a site of reduced subjectivity.[31] Just as the speaker is robbed of a means of communication, the expressive range of the music is stunted, locking into a groove. The hypnotic effect of the music is aided by the fact that the listener doesn't have semantic meaning to attune to. I hear Holiday's upward inflection as indicating some sort of exclamatory gesture, although I must rely on the pragmatics of her delivery rather than semantics to arrive at that conclusion.

At 10:21, we hear Holiday's voice start to be reassembled into semantically meaningful statements—«Jesus Christ, no!»—but even as semantic content is reintroduced, it's clear that the

[31] According to Rebecca Leydon, "Music that confounds hierarchic listening altogether because of a preponderance of undifferentiated 'riffs' may suggest a 'will-less' or 'automatized' subject." Rebecca Leydon, "Towards a Typology of Minimalist Tropes," *Music Theory Online* 8, no. 4 (2002), http://www.mtosmt.org/issues/mto.02.8.4/mto.02.8.4.leydon.html, [14]. Rebecca M. Doran Eaton, "Marking Minimalism: Minimal Music as a Sign of Machines and Mathematics in Multimedia," *Music and the Moving Image* 7, no. 1 (2014): 3–23, links the minimalist aesthetic with the mechanical in the use of minimalist music in commercials, films, and operas.

pragmatic content, the exclamatory nature of the statement, is more important to its place in the music than its meaning. Finally, at 10:53, in the final 20 seconds of the piece, TV restores the full power of speech to Holiday's voice. Here she once again speaks full sentences, unaccompanied except for hits on the downbeat from the orchestra. By allowing the spoken voice alone to be the final thing that the listener hears, I interpret this final vocal solo as TV's means of returning Holiday's speaking voice to its intended purpose, spoken communication, after 11 minutes of using it for musical purposes.

In the moments where TV strips Holiday's voice of its ability to communicate semantically, the inequity between the speaker and the composer in this virtual, intermundane collaboration is clear. The speaker's voice is subservient to the composer's whims. In the case of "Billie," the use of a Black woman's voice by a white male composer adds a more complex layer to this power imbalance. Scholars have identified a similar racial power disparity in Steve Reich's *It's Gonna Rain* and *Come Out*, and in particular I think the work of George Adams and Maarten Beirens can be applied to the interaction of TV and Holiday in "Billie." Adams sees Reich's aforementioned tape works as examples of what Marie Thompson calls "white aurality" in Eurological sound art, the muffling and erasure of the sociality and other sonic practice in favor of a "sound-itself."[32] Likewise, when TV reduces the sound of Holiday's voice to merely an object for musical creation, it stifles Holiday's subjective experience. Beirens, regarding the gradual breakdown of semantic content in Reich's use of tape loops, writes,

> "The process of removing the recognisably human aspect of the voice serves to heighten the perception of the machine as driving forward this transformational process. Stripping away the personal layers—the individuality of the speaker,

[32] George Adams, "Listening to Conceptual Music: Technology, Culture, Analysis" (Chicago, University of Chicago, 2019), 219–22. See also Marie Thompson, "Whiteness and the Ontological Turn in Sound Studies," *Parallax* 23, no. 3 [84] (2017): 274.

> which disappears together with the recognisability of his or her voice—brings about the notion of the *impersonal*. [emphasis Beirens's][33]

In a similar way, when TV breaks Holiday's voice down into phonemes, he draws our attention away from Holiday as a subject with a body and instead emphasizes the technologically mediated process by which he alters the recorded voice, highlighting the object-ness of the speech recording. While other parts of the piece showcase the influence of Holiday's voice on the music, this particular musical moment overpowers Holiday as an individual and turns her speech from an embodied communicative medium to an abstract sonic vehicle in service of TV's composition.

*****

"Billie" proves a case study in how expressivity in music and in speech can be mapped enhance or change meaning. We have seen a variety of ways in which musical expressivity is mapped onto the voice. The use of the same speech segment in the **A** and **B** sections shows us the power that music has to inform meaning in connection with a text. By setting a single spoken phrase to different musical moods and gestures (and by changing the speech surrounding this spoken motive), TV encourages the listener to take away different meanings from the same semantic content. The juxtaposition of Holiday's speaking and singing voice stresses the (at times blurry) line between these two means of communication. The use of timbral sonic analogs (produced both by the human performer mediated via musical instrument and by digital processing) highlights a feature of Holiday's later-life vocal expression and in doing so focuses on Holiday as an individual by emphasizing her deteriorating body. TV's curation of sonic space through processing techniques highlights the expressive range of Holiday's voice. The curation of speech objects and sonic space by TV results in reorderings that change linguistic meaning and

[33] Maarten Beirens, "Voices, Violence and Meaning: Transformations of Speech Samples in Works by David Byrne, Brian Eno and Steve Reich," *Contemporary Music Review* 33, no. 2 (2014): 219.

allow musical meaning to fill space left when semantics are abandoned. Most importantly, by highlighting features of vocal expression through musical means, the composer calls attention to subject position of the speaker. "Billie" glorifies Holiday's expressive range, the musicality of her voice, but also demonstrates the power that the composer exerts in works that musicalize speech by changing and enhancing meaning.

But at the same time, we can see the underlying power imbalance between the composer and speaker at work; TV has made authorial interventions into Holiday's delivery through technological means. It also highlights powers imbalance between TV and Holiday along the lines of race and gender, bringing up ethical concerns in the use of the recorded speaking voice in music. These concerns will be taken up later on in this chapter as I discuss speaker agency in virtual collaboration.

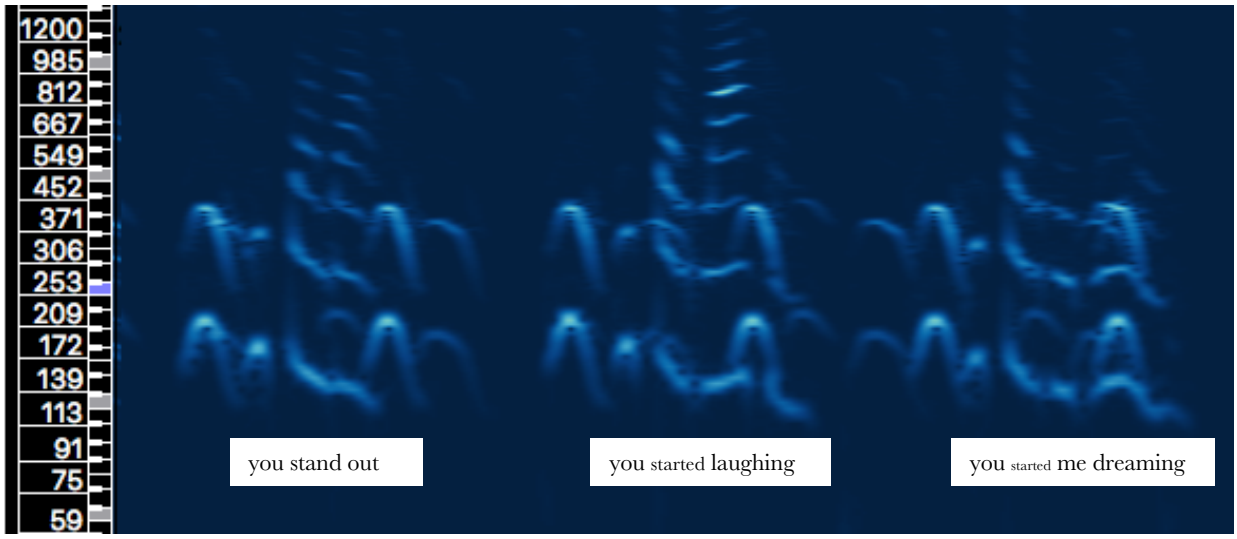
DJ Spooky, "Perpetual/Pop Titles 'You'"

In 2008, artist and experimental hip-hop producer DJ Spooky that Subliminal Kid released *Sound Unbound: Excerpts and Allegories from the Sub Rosa Archives*. The album is comprised remixes created from material released by Belgian record label Sub Rosa, which specializes in avant garde, electronic, and noise music. DJ Spooky describes this album as an "audio essay" that "foregrounds some issues that I think the 21st century really needs to understand—art isn't about objects anymore. Art is a continuous investigation into humanity's place in the world."[34] The album features many speaking voices set to music, treading a line between the recorded voice as object, a sample, and as subject, deriving from a human source. The release of the album coincided with the publication through MIT Press of the collection of essays *Sound*

[34] "DJ Spooky (That Subliminal Kid)," Sub Rosa, accessed December 12, 2019, https://www.subrosa.net/en/catalogue/soundworks/dj-spooky-that-subliminal-kid_01.html.

Unbound: Sampling Digital Music and Culture, edited by DJ Spooky himself (under his legal name Paul D. Miller).[35] The collection focuses on sampling's place in 21st-century art and music, and comes with a copy of the *Sound Unbound* album as a companion. I'm interested here in the way DJ Spooky samples Pamela Z's *Pop Titles "You"* in the track "Perpetual/Pop Titles 'You,'" particularly in how the musical setting accompanies and heightens Z's vocal expression.

Figure 4.7. Melodic frequency spectrogram, Pamela Z, *Pop Titles "You"*, 0:12–0:18

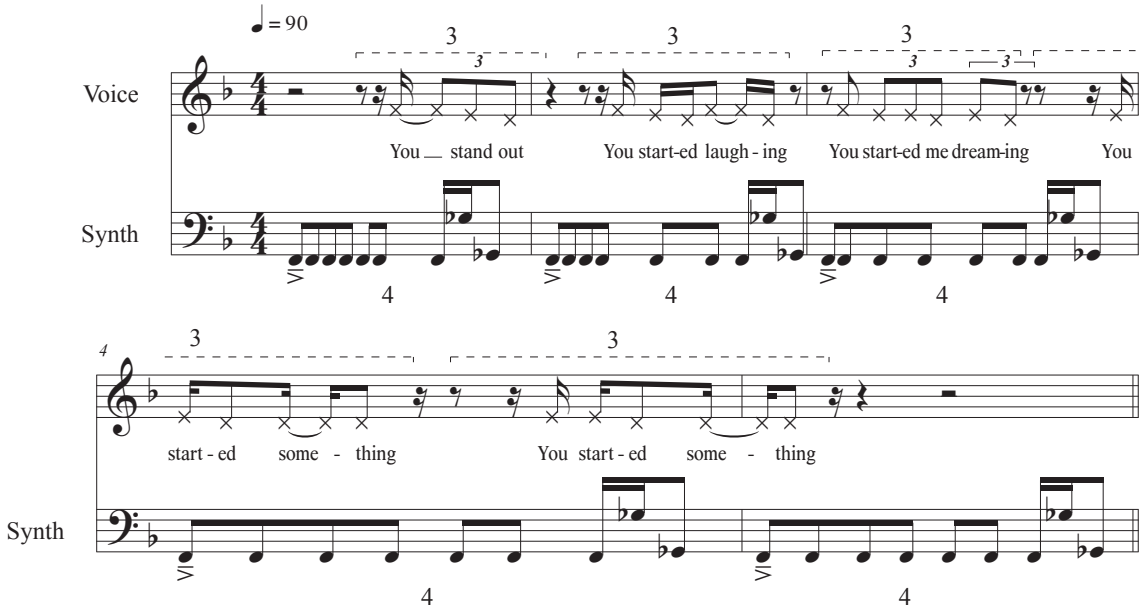


Unlike the wide expressive range of Billie Holiday's voice in "Billie," Z's vocal expression is rather muted in *Pop Titles "You."* At the heart of this piece is a list of song titles read by Z in an unbroken manner. (See my analysis in Chapter 2 for a more detailed overview of this work.) While Z's voice does carry a sort of melodious quality, the semantic content of most of the spoken phrases she delivers does not carry a particularly strong emotional charge. Z's voice is expressively understated—more a recitation than an expressive reading—and stands out in comparison to her penchant for experimenting with extended vocal techniques. Figure 4.7 shows

[35] Paul D. (DJ Spooky) Miller, ed., *Sound Unbound: Sampling Digital Music and Culture* (Cambridge, Mass.: MIT Press, 2008).

a spectrogram of the first three spoken phrases from this piece. We can see and hear that Z's emotionally understated voice follows a falling contour, one that is quite similar in pitch space and time regardless of the words being said. As discussed in Chapter 2, the repetitive nature of Z's speech phrases has a sort of hypnotic effect. The evenness of her delivery sets up an expectation that all phrases will be delivered in largely the same way.

Figure 4.8. Transcription of first five spoken phrases, DJ Spooky, "Perpetual/Pop Titles 'You'", 0:22–0:32



DJ Spooky's musical setting in "Perpetual/Pop Titles 'You'" reflects this highly repetitive, minimal structure of the spoken phrase by likewise employing repetitive, minimal musical phrases. Figure 4.8 gives an example of the musical structure taken from the first five spoken phrases (0:22–0:32). The music is driven by an ostinato bass synth line, one measure in length in 4/4 at 90 bpm. "Perpetual" has very simple pitch landscape, just F2 and neighboring G-flat2 (the latter emphasized, however, by octave leaps to G-flat3). This clear pitch center on F perceptually reworks Z's speech. Although her voice is still clearly construed as speech rather than song, the tonal context of the track serves to sharpen the edges, so to speak, of the pitch material of Z's

spoken phrases. I've used crossed note heads in my transcription to represent the impreciseness of these pitches, but nonetheless, one can hear each phrase occupy a clear range of a minor third from F4 to D4, serving to emphasize the tonic.

On the one hand, the repetitive nature of the materials used to establish tonal and temporal frameworks serves to underscore the droning quality of Z's voice, mimicking her repetition and highlighting it by casting the pitch content of her speech into a clear tonal context. But on the other hand, DJ Spooky introduces tension by setting Z's speech to music that proceeds at a different tempo to that of Z's original piece. Whereas the repeated «you»s of *Pop Titles "You"* occur once per second and form the tactus for the work, the tactus of the prominent accompanimental synth line in "Perpetual/Pop Titles 'You'" occurs at 90 bpm. This sets up a tension between the metrical framework established by Z's composition and that suggested by DJ Spooky's synth bass. As we can see in the transcription in Figure 4.8, the first spoken phrase of begins in an unexpected and metrically-dissonant position, the anacrusis of beat 4. It is almost immediately apparent that the phrasing of the voice simply does not fit the 4/4 metrical framework, and later phrases confirm this unsettled marriage of speech rhythm and musical meter. Each of the similarly timed spoken phrases, which had fit neatly into a measure in Z's piece, occurs over the span of three beats in the 4/4 time established by DJ Spooky's synth bass. What was regular in Z's work now pulls against the established meter of "Perpetual." This grouping dissonance persists through much of the track, creating an underlying tension between the repetitive layers of speech and musical accompaniment.[36]

[36] On the notion of grouping dissonance, see Harald Krebs, *Fantasy Pieces: Metrical Dissonance in the Music of Robert Schumann* (New York: Oxford University Press, 1999). Krebs abbreviates the kind of grouping dissonance that is evident in DJ Spooky's "Perpetual" as G4/3 to show that the underlying groupings of four beats are disrupted by overlaid groupings of three beats.

This tension, however, is one that is familiar to listeners of this genre—electronic dance music—which commonly features rhythmic dissonance laid over short configurations of tonal bass line and percussion repeat to form a groove in 4/4.[37] In this case the grouping dissonance crafted by DJ Spooky between spoken and musical phrases is normative, reinforcing the unmarked quality created by Z's unobtrusive vocal expression and the musical accompaniment that is repetitive and not conventionally "expressive." All of these factors combined yield an expressively understated musical work, a sharp contrast to the wild mood changes and highly expressive voice of Jacob TV's "Billie."

Recall that Jacob TV used substantial audio processing on Billie Holiday's voice. In DJ Spooky's piece, however, Z's voice is presented as she recorded it: a fairly natural rendering of the voice through the recorded medium. It is thus an example of what Miriama Young calls a "pure" voice, a recorded voice cultivated to sound "full, rich, crisp, and present."[38] It should be noted that DJ Spooky leaves Z's original composition largely untouched, simply cutting it off when the track ends. Z's piece is presented as-is, and DJ Spooky builds his musical surface around it. An argument could be made that by using Z's musical work and her voice as she crafted them, DJ Spooky gives Z an elevated role in the collaboration. DJ Spooky's musical setting, even though it produces metric tension, seems designed to underscore the restrained expressivity of Z's original piece. His addition of regular repeating pitch and rhythmic frameworks serves to reinforce the minimal nature of Z's spoken phrases, her words, and her own

---

[37] Mark J. Butler, *Unlocking the Groove: Rhythm, Meter, and Musical Design in Electronic Dance Music* (Bloomington: Indiana University Press, 2006), 5–6.

[38] Miriama Young, *Singing the Body Electric: The Human Voice and Sound Technology* (Farnham, Surrey: Ashgate, 2015), 131 makes a distinction between the "digitally 'pure' vocal presence," in which the recorded voice appears to be clean despite subjection to make post-production processing techniques, and the mediated voice, which is processed in such a way that its alteration is obvious to the listener.

musical structures. It's important to note, however, that DJ Spooky, like Jacob TV, is using a woman's voice (and her pre-existing cultural product) to create a new work over which he has ownership.

As in "Billie," we can hear the musical accompaniment as an enhancement of the voice's meanings, both semantic and pragmatic. This added layer of musical meaning reinforces the voice while subtly urging the listener to also reinterpret the voice. DJ Spooky's minimal musical frameworks bring out the repetitive nature of Z's spoken phrases and the limited range of expression in her voice's pitch contour and rhythm. By doing little to alter Z's voice, DJ Spooky acknowledges the subject, the human, the musician and artist, behind the recorded voice.

Virtual collaboration and speaker agency

Other scholars who have considered recorded speech in music have largely taken a negative stance on its connection with human subjects. Steven Connor views the recorded voice in postmodern music as decomposed and degraded, incapable of recapturing its former humanity.[39] David Code argues for a loss of identity that the speaker undergoes in Paul Lansky's *Six Fantasies on a Poem by Thomas Campion* (1978–79) as the intelligibility of the speech is lost in the first, third, and fifth movements.[40] As noted earlier, scholars concerned about the extent to which recorded speech retains its human elements have often focused on Steve Reich's tape works for voice, *It's Gonna Rain* (1965) and *Come Out* (1966). Maarten Beirens, for example, sees a violence in the degradation of the voice in these works, which strip the voices of African American speakers of their semantic elements until they become purely sonic objects.[41] Martin Scherzinger, for his

[39] Steven Connor, "The Decomposing Voice of Postmodern Music," *New Literary History* 32, no. 3 (2001): 479–80.
[40] David Loberg Code, "Observations in the Art of Speech: Paul Lansky's Six Fantasies," *Perspectives of New Music* 28, no. 1 (1990): 167.
[41] Beirens, "Voices, Violence and Meaning," 217.

part, hears *It's Gonna Rain* as "gradually transform[ing] the voice of a black man into animal sound."[42]

Similar arguments can be levied against TV in "Billie": a white composer borne of the Western concert tradition uses the voice of a Black woman to his own compositional ends, robbing her of agency, particularly as her voice is divested of semantic meaning in the final section of the piece.[43] TV treats Holiday's voice as a sort of embodiment of her life story along the lines that Nina Sun Eidsheim identifies in so many critical writings on Holiday's voice. As Eidsheim warns us, "reducing Holiday's voice to an a priori—that is, limiting perspectives on her voice to the projection of an idea of her life—reduces her subjectivity and artistic agency to the oft-narrated arch of her biography."[44] In TV's "Billie," we hear an *interpretation* of Holiday's voice, shuffled and pasted together in a way that signifies Billie Holiday and whatever we may think her voice tells us about her life, race, or gender, but each of TV's interventions further remove us from the living human subject who projected this voice.

I see TV's use of Holiday's voice as exhibiting both the kind of appropriation and violence heard in Reich's tape works *and* a reverence for Holiday as a musical cultural figure, a desire to showcase the uniquely melodic quality of her speaking voice. Whereas the voices that Reich uses—those of Daniel Hamm and Brother Walter—come from speakers who are relatively

---

[42] Martin Scherzinger, "Curious Intersections, Uncommon Magic: Steve Reich's It's Gonna Rain," *Current Musicology* 79/80 (2005): 217.

[43] For more on the how the musical voice is shaped by cultural forces that reinforce existing power structures of race and gender, see Nina Sun Eidsheim, *The Race of Sound: Listening, Timbre, and Vocality in African American Music* (Durham: Duke University Press, 2019). On how Western sound art muffles the sociality of sound—particularly with regards to race—under the guise of "sound-itself," see Thompson, "Whiteness and the Ontological Turn in Sound Studies," 266–82. On the "sonic color line," the socially constructed boundary that enables us to code sounds as belonging to a race and therefore discriminate by assigning sounds differential cultural, social, and political value, see Jennifer Lynn Stoever, *The Sonic Color Line: Race and the Cultural Politics of Listening* (New York: New York University Press, 2016).

[44] Eidsheim, *The Race of Sound*, 159.

unknown, TV uses the voice of a popular figure considerably more famous than himself. There's a cultural cachet to Holiday's voice that TV capitalizes on, and Holiday's voice is almost certainly going to be more familiar to the listener than TV's "composer's voice." After all, as Beirens points out,

> Because of the use of a specific recording, it becomes impossible to separate what is being said from the person who is saying it. If only through the idiosyncrasies of speech and pronunciation, these words remain inextricably bound to the speaker, who is audibly (though not physically) present. Borrowing such instances of recorded speech seems impossible without disclosing the 'identity' of the speaker.[45]

He goes on to argue that the implied presences of the speaker, along with the musical structure, text, and voice, make possible the communication of actions and emotions in works for recorded speech.[46] Because of the ubiquity of Holiday as a public figure, it is arguably her voice rather than TV's that is foregrounded in "Billie."

"Billie" acknowledges Holiday's subject position as a singer by treating not only Holiday's speaking voice but her singing voice as well, using a recorded outtake. The outtake of "End of a Love Affair" that TV uses was released posthumously, and when taken as a musical work in its own right I find it jarring in its sparsity; when I listen I find myself wondering if a living Holiday would have consented to its release to the public. Both this outtake and TV's employment of Holiday's voice bring to the fore how ethically fraught the use and appropriation of the posthumous recorded voice can be.

I want to consider now what ethical obligations a composers might have to speakers whose voices they are using in the creation of a musical work. The agential status of composers and speakers indeed creates a possibility for inequitable treatment along identity lines such as race or gender, particularly because the speaker may be an unknowing participant in the

[45] Beirens, "Voices, Violence and Meaning," 212.
[46] Ibid.

compositional process and therefore already has limited agency. As I will discuss in this section, I see varying levels of equitability in such collaborations based on the level of involvement of the speaker and the treatment of the voice by the composer, but throughout this discussion I wish to treat the speaker as a human actor rather than a sound object.

As does Jacob TV in "Billie," DJ Spooky uses the voice of another musician in "Perpetual/Pop Titles 'You'"—with the added complication that Pamela Z's voice comes from a complete musical work itself. Whereas Holiday might have imagined her vocal take from "End of a Love Affair" as one orchestrational part of a musical work that included instrumentation, Z's *Pop Titles "You"* stands on its own as a closed tape work, fully-realized. DJ Spooky does little to intervene in or alter Z's voice, presenting it largely as she composed it in *Pop Titles "You."* Z as a composer is heavily invested in exploring the musicality of speech; I read DJ Spooky's remix, which brings out emergent musical features of Z's speech in its recontextualization of her voice, as an homage to her artistic interest in speech as music. DJ Spooky's virtual collaboration with Z, while still uneven, seems more equitable than Steve Reich's with Daniel Hamm and Brother Walter or Jacob TV's with Billie Holiday, in which TV appropriates Holiday's voice to make a new artwork that gains distinctiveness through (1) the cultural cache of Holiday and (2) TV's extreme reconfiguration of Holiday's voice.

I argue that by using recordings of the human voice, composers are trying (whether successfully or not) to engage with the humanity of the speaker—albeit a humanity that has been radically transformed and complicated by its sonic virtual manifestation. In her work on late modernist music, Marcelle Pierson contends that the reduction of the voice to a sound object still leaves residual meaning, be it melody, semantic information, or persona.[47] Following Pierson

[47] Marcelle Coulter Pierson, "The Voice under Erasure: Singing, Melody and Expression in Late Modernist Music" (University of Chicago, 2015), 89.

then, the speaker's subjectivity is not entirely lost in these musical works, although the nature of the subjectivity becomes altered when its voice is stripped from its body. But on the flipside, the composer's voice, which accounts for the implicit musical persona of a piece of music (per Edward Cone), is likewise limited when it yields some level of compositional control to the speaking voice.[48]

Why should the composer in this era (late 1960s to the present) want to relinquish some level of authorial control? According to Jonathan Sterne, "the great dilemma of our musical culture today is the position of the composer, who is an isolated figure, cut off from the vast majority of the community, sending out his messages into the void and wondering if anyone is listening, condemned always to speak to an essentially passive audience, with whom the closest relationship he can hope to achieve is that of producer to paying customers at a concert."[49] But as Stanyek and Piekut, Jennifer Iverson, and Eric Clarke and Mark Doffman have shown, technological mediation in recent music is often a collaborative process.[50] I posit then that the recorded voice presents a means to pick apart the lone-composer narrative—to loosen the compositional control exerted by the composer, who must learn to adapt the pre-determined pitch, rhythmic, and phrasing materials presented by the recorded voice. By ceding some level of

---

[48] Edward T. Cone, *The Composer's Voice* (Berkeley: University of California Press, 1974), 25f.
[49] Jonathan Sterne, "On the Future of Music," in *Cybersounds: Essays on Virtual Music Culture*, ed. Michael D. Ayers (New York: Peter Lang, 2006), 257.
[50] Stanyek and Piekut, "Deadness," 304–324 consider the effect of the performers and recording engineers as well as the limitations and affordances of recording technology (as created and used by human actors) in the collaboration between deceased Nat King Cole and his living daughter Natalie Cole on the track "Unforgettable" (1991). Jennifer Iverson, "Invisible Collaboration: The Dawn and Evolution of Elektronische Musik," *Music Theory Spectrum* 39, no. 2 (2017): 200–222 examines the early output of the Westdeutscher Rundfunk as the result of a collaborative network of composers, machines, technicians, and psychoacousticians. For more on the potential for collaboration in contemporary music, see Eric F. Clarke and Mark Doffman, *Distributed Creativity: Collaboration and Improvisation in Contemporary Music* (New York, NY: Oxford University Press, 2017).

authorship to the speaker, the composer elevates the person behind the voice to the level of collaborator. At the same time, the voice is reconfigured along the lines of one of Frederic Jameson's tenets of the postmodern experience, the "waning of affect," wherein the idea of the bounded self begins to fray.[51] The musical work results from the marriage of the recorded speaking voice and the composer's authorial voice. As I have already noted, however, the collaborative aspect is not a give-and-take between equal partners. The recorded voice of a human subject is treated as a compositional *object*. It is unable to give more than its fixed utterances, presented as they were recorded or manipulated by the composer to transcend their original temporal and pitch boundaries. But that does not mean that the presence of the speaker-as-subject is not felt in these works.

These compositional objects made of speech are based are recordings of the past, often of now-dead speakers, and typically do not allow for direct interface between the speaker and the listener. There are, however, some pieces that involve collaboration between composers and living speakers, who have varying degrees of influence over the musical work. In *Six Fantasies on a Poem by Thomas Campion* (1978–79), Paul Lansky asked his wife Hannah Mackay to record her voice. All sounds in the resulting piece derive from a single take, however, indicating that Mackay's involvement was rather limited. Furthermore, while the voice that Lansky uses comes from a speaker who is alive and close to him, he and Mackay are using the *words* of long-dead Thomas Campion (1567–1620).

Much of Pamela Z's output for recorded speech likewise presents a collaboration between living speakers who have real-life interaction with the composer. *Ways of Looking* (2019) is built around the voices of the performing group Eighth Blackbird, who commissioned the piece.

[51] Perry Anderson, *The Origins of Postmodernity* (London: Verso, 1998), 57.

Geekspeak (1995) uses recordings from interviews with software engineers with whom Z came into contact during an artist residency at Xerox's Palo Alto Research Center. The interviewees were aware of Z's plans to turn their responses into a musical piece, even if they did not have direct compositional input. For composers like Lansky and Z, although the composer has final authorial control, he or she has knowingly invited others into the compositional process.

In cases in which the speaker is dead, this does not completely rule out the possibility of imagining these works as a collaboration between the composer and the speaker. The speaking voice is heard as emanating from a body, an individual. Even in Peter Ablinger's *Quadraturen III,* in which the voice is mediated by instrumental speech synthesis rather than a speech recording, the player piano at least has the potential to be read as an extension of the original speaker's body if it is attended to by the listener as comprehensible speech. The speech object "has an effect"—one of Stanyek and Piekut's requirements for agency—by conveying meaning to a listener. The resulting musical work represents a collaboration between the speaker, who lends his/her voice (including words, accent, emphasis, expression, etc.) and the composer, who provides a means of mediating the voice, choosing the parameters of the temporal and pitch musical framework and translating the acoustical information of the speaking voice to the medium of music.

Furthermore, the speaker is treated as a performer. It is the vocal persona of Edward Cone's triad of musical personas (vocal, instrumental, and composer), the only incarnate persona that expresses itself through the human voice.[52] The voice provides an embodied focal point to which the listening subject can relate. The voice is indeed more readily recognizable as a subject than a composer's implicit persona. But beyond merely being the embodied voice of these

[52] Cone, *The Composer's Voice*, 18.

musical works, the speaker-as-performer, as George Adams argues, is "an agential force expressed through the act of singing or playing music" separate from the sounding voice.[53] The performer's voice "foreground[s] her interpretive contributions to the composer's musical work."[54] Although in composing with the recorded speaking voice the act of interpretation is done by the composer, we nonetheless hear the *expressive* power of the speaker-as-performer and her musicalized voice.

Musical works that use the recorded voice as a compositional object are a site of tension between modernist and postmodern approaches to art, an attempt to recapture the self in the postmodern world. At first glance, this corpus of music has several hallmarks of postmodern art. There is the erasure of the boundary between high and low art—consider how many disparate genres of music it appears in, from YouTube video parodies to institutional contemporary composition. It also uses the quotation of a speaking voice as the substance of the artwork.[55] Michael Klein argues that by using quotation in art, the personal style of the composer or author is lost except in the arrangement of the quotations.[56] But although this corpus relies on borrowing voices and linguistic content from its speakers, it *does not* try to hide the author and his or her style. Instead, it performs a recontextualization—a musicalization—of the speaker set within an author's style. The speaker's subject is obscured by the composer, and not the other way around; The speaking voice is often reordered and digitally manipulated in service of a musical style.

In musical works that are organized around the recording speaking voice, the linguistic meanings conveyed by the speaker—be they semantic, phonological, or pragmatic—become

---

[53] Adams, "Listening to Conceptual Music," 168.
[54] Ibid.
[55] For a discussion of the qualities of postmodern art, see Fredric Jameson, *Postmodernism, or, The Cultural Logic of Late Capitalism* (Durham: Duke University Press, 1991), 62–66.
[56] Michael Leslie Klein, *Music and the Crises of the Modern Subject*, Musical Meaning and Interpretation (Bloomington, Indiana: Indiana University Press, 2015), 135–38.

inextricably linked to musical meaning, forming the basis for (or at least informing) the musical structure. Musical expressivity interacts with and underscores spoken vocal expression. The speaker becomes an indelible part of the compositional process, even if he/she has no direct collaborative input. Musical works in this oeuvre present a uniquely modern assemblage of actors linked together by technological means.

Chapter 5: Musicalized Speech in Stylistic Frameworks

In the first part of this dissertation I discussed musical works from wildly differing genres and epochs concurrently for the sake of identifying trends in compositional approach across a diverse corpus of music. I've thus far been using the term "musical framework" fairly broadly to describe any sort of pitch and temporal structure that orders musical sound. Recorded speech is musicalized through the imposition of musical frameworks of pitch and time. But moreover, the structures of the musical frameworks differ in each work due to the composer's individual approach to composition, the genre in which she is writing, and the composer's influences as they manifest musically. Simply put, all of these factors constrain the basis for compositional choice, meaning that the possibilities for musical structure are limited by them—in other words, they form the basis for the work's musical style, broadly defined.

In this chapter, I take a deeper look at the musical scaffolding in individual works that grounds recorded speech in order to make that speech sound musical. To do this, I consider how a composer's musical influences and personal compositional strategy affect the formation of musical frameworks. These unique frameworks, constructed for individual musical works by individual composers, are what I call *stylistic frameworks*. Each musical work will have a distinctive stylistic framework of pitch, time, timbre, orchestration, etc. that is decided by the composer's compositional strategy, genre, and outside influences like training or personal taste.

I envision the musicalization of recorded speech as an act of musical analysis done to the speech recording by the composer and then overlaid and sculpted onto the chosen stylistic frameworks. The stylistic frameworks that a composer uses in these works are not necessarily derived from the sonic structure of the recorded speech itself. Instead, they are a means of interpreting the speech musically, a lens through which the composer tries to make sense of the

phonological features of speech. These stylistic frameworks bring with them a set of constraints and expectations for shaping sound, and serve as a means to limit the number of possibilities for musicalizing speech. But the force of speech sounds and stylistic frameworks of music act in a reciprocal relationship in the composition of these works—the phonological features of the individual speech recordings likewise limit and influence the possibilities for musicalization.

The main goals of the chapter are twofold. First, I examine more deeply the compositional choices that shape the musicalization of speech. It's by no means a given that recorded speech can be turned into music, and composers who experiment with recorded speech do so in wildly varying ways that musically communicate different things to their audience. Second, I consider composers as agents who have an effect on the interpretation of the recorded human voice, and the analyses I offer in this chapter provide an opportunity to further explore the power play between the speaker's voice and the composer's voice that I postulated at the beginning of Part II.

I will do this with two case studies. First, I will approach Charles Dodge's *The Story of Our Lives* (1973–74) as an interrogation of the line between speech and song in recitative. Dodge uses musically pitched speech synthesis to create a dramatic narrative between a man and a woman who join in conversation with a book about their lives that threatens to destroy their happiness. Because recitative has long been held as a sort of musical analog for speech, I will explore which features of recitative as a musical genre are retained when such music is confronted with genuine speech and how Dodge sets his singing speech synthesis apart from speech itself. Second, I will return to Peter Ablinger's "Arnold Schoenberg," a movement from *Voices and Piano* (begun 1997, ongoing), as a musical work for recorded speech that serves as an homage to the influential 20th-century composer whose voice is used as a compositional object. In works like this, stylistic allusion conveys referential meaning by evoking extra-musical associations with the speaker's

music.[1] Hence the composer is able to interact with the speaker as a fellow composer, the speaker as a historical figure, and, finally, the speaker as a human subject with a voice.

Stylistic frameworks

At the heart of this dissertation has been the thesis that while both speech and music are systems for organizing sound, the specific syntactical rules and expectations of each put them in conflict when placed together. The grammatical and syntactic features of music, however, are much more loosely held than those of natural language. In Part I of the dissertation I provided a brief overview of the ways music organizes sound differently than language (more reliance on repetition, hierarchically organized meter, discrete pitches). Musical frameworks that organize the disposition of temporal onsets or pitch relationships are essential to our understanding of musical gestures because they make it possible to set up larger patterns of expectation for the listener.[2] Simply put, they form the basis for interpreting a particular musical work or passage—they shape the musical grammar. It is important to note that the exact structure of musical frameworks varies from piece to piece although composers may participate in a tradition of shared compositional strategies for developing and deploying these frameworks. Each piece has a unique confluence of its composer's compositional strategies and influences and stylistic expectations, and this confluence forms the basis for the musical structure.

Going forward, I will explore this confluence in terms of *stylistic frameworks*, a name I give to musical frameworks as they are uniquely constructed in individual musical works. But first we

---

[1] For more on stylistic allusion, see Peter Manuel, "Music as Symbol, Music as Simulacrum: Postmodern, Pre-Modern, and Modern Aesthetics in Subcultural Popular Musics," *Popular Music* 14, no. 2 (1995): 231.

[2] See Lawrence Michael Zbikowski, *Foundations of Musical Grammar*, Oxford Studies in Music Theory (New York, NY: Oxford University Press, 2017), 124–126 on the importance of temporal and tonal frameworks to understanding musical gestures.

should consider the somewhat amorphous concept of "style." Leonard Meyer defines it, in quite broad terms, as "a replication of patterning, whether in human behavior or in the artifacts produced by human behavior, that results from a series of choices made within some set of constraints."[3] Speaking more directly about music, Jan LaRue asserts that the style of a piece consists of the predominant choices of elements and procedures a composer makes in developing movement and shape (or perhaps, more recently, denying movement or shape). [4] By extension, we can perceive a distinguishing style in a group of pieces from the recurrent use of similar choices. A composer's style as a whole can be described in terms of a confluence of both consistent and changing preferences in her use of musical elements and procedures.

Perhaps the best definition of style for my purposes comes from Mariateresa Storino and her colleagues, who argue for a model of style analysis based on grammatical rules, or "the idea that music is the result of physical and intellectual constraints aimed at selecting sounds and organizing sound sequences of human communication by means of rules."[5] Stylistic frameworks are a manifestation of the limiting of musical possibilities based on the constraints like genre and personal compositional strategies.

"Style" can exist on various hierarchic levels: constraints of a whole culture, of an epoch or movement, of an individual composer, or of a single work of art.[6] I will be focusing on musical style within the two middle levels of this hierarchy—the levels of movement (or genre) and composer—and how it exerts force on a single work of art. But the highest and lowest levels are not entirely absent from this discussion. As I noted at the end of Chapter 3, the use of recorded

---

[3] Leonard B. Meyer, *Style and Music: Theory, History, and Ideology* (Philadelphia: University of Pennsylvania Press, 1989), 3.
[4] Jan LaRue, *Guidelines for Style Analysis*, 2nd ed. (Warren, Mich.: Harmonie Park Press, 1992), ix.
[5] Mariateresa Storino, Rossana Dalmonte, and Mario Baroni, "An Investigation on the Perception of Musical Style," *Music Perception: An Interdisciplinary Journal* 24, no. 5 (2007): 418.
[6] Meyer, *Style and Music*, 13.

speech in the structure of music is to a certain extent cultural. It appears in different genres and art movements, high and low, across Western music from the last roughly half-century. On the other end of hierarchy, the use of recorded speech in composition is at times simply a novel approach by composers: a number of the composers I have discussed have made these speech works as sort of one-off experiments, anomalies in their larger oeuvre. The sonic structure of individual works themselves are a convergence of musical style and the phonological and expressive features located in the speaker's voice.

In this chapter, I will consider both genre and individual compositional style as frameworks for the musical interpretation of speech because I believe that these two hierarchical levels have the most direct influence on the structure of the resulting work and can account for how the compositional strategies that I identified in Part I are utilized. We've thus far seen a wide variety of styles used in this particular approach to composition. If made to compose a piece of music from the same recording of speech, Steve Reich, Jeff Richmond, and James Tenney('s ghost) would assuredly produce wildly different soundworlds. Based on what we've heard in their oeuvre, both Reich and Richmond might compose something tonal with a strong metrical profile, although their deployment of those broad musical frameworks would not sound terribly similar. Instead, they would likely fall into the genre expectations of minimalism and pop music, respectively. Were we to come across a previously undiscovered work by Tenney that used recorded speech, we might expect it to be microtonal and to leave the listener with a much looser sense of hierarchically organized musical time. Some of these differences would come down to individual style, and some would be features of the traditions of shared compositional strategies in which each composer is working. These musical frameworks—more specifically, these stylistic frameworks—provide a lens through which the composer interprets the sounds of speech.

I see style as the basis for the overlay of musical gestures onto recorded speech. The musical frameworks that a composer uses in these works are not necessarily derived from the sonic structure of the recorded speech itself. Instead, they act as a filter through which the composer tries to make musical sense of the phonological features of speech, a means of interpreting the speech musically. Style brings with it a set of constraints and expectations, and serves as a way to limit the number of possibilities for musicalizing speech. I see stylistic frameworks as a set of specialized musical frameworks that a composer uses to make choices about the mapping of music onto speech.

Charles Dodge's *The Story of Our Lives* and the line between speech and song in recitative

I'll begin by exploring the influence of genre conventions on the implementation of stylistic frameworks. We've seen an example of this in Chapter 4 in DJ Spooky's "Perpetual/Pop Titles 'You,'" which remixes Pamela Z's *Pop Titles "You"*. DJ Spooky adds an instrumental accompaniment of synth and drum machine that conforms to expectations for electronic dance music: 4/4 meter and a tonal pitch framework. When placed against this accompaniment, Z's spoken phrases set up rhythmic dissonance, another common feature of the genre.[7] The addition of a pitch framework emphasizes and draws out the pitch content of Z's voice, making the repeating descending contour of her voice more salient. Through these means DJ Spooky's musical accompaniment reinterprets or enhances phonological features of Z's voice, bringing out their potential for musicality. Stylistic frameworks are thus a means of setting up musical expectations through which listeners can hear speech as musical or through which composers can alter speech to suit the musical structure they create.

[7] Mark J. Butler, *Unlocking the Groove: Rhythm, Meter, and Musical Design in Electronic Dance Music* (Bloomington: Indiana University Press, 2006), 5–6. Butler discusses genre expectations for electronic dance music in his introduction.

To further explore how composers employ stylistic frameworks that can serve as a frame of reference by which to musicalize speech, I'll examine Charles Dodge's *The Story of Our Lives*, a computer synthesis piece that uses speech analysis and synthesis to convert plain speech into what Dodge calls a "dialogue in recitative."[8] I propose that this piece is an interrogation of the line between speech and song in recitative, a genre that has long been held as a sort of musical analog for speech in the absence of speech itself—a sort of heightened artistic interpretation of speech sounds. The question, then, is which features of speech are transferred onto recitative and how are they transformed into music? I will explore which features of recitative as a musical genre Dodge retains as he musicalizes actual speech and how Dodge sets his singing speech synthesis apart from speech itself. In doing so, I explore the ways in which generic and individual composer's stylistic frameworks are used to musicalize the voice works for recorded speech.

Dodge's compositional approach is outlined in the liner notes for his album *Synthesized Voices*:

> The technique used to create the synthetic voices is called speech synthesis-by-analysis. For this method of speech synthesis, only those words, phrases and sentences which have been spoken into the computer previously (via an analogue-to-digital converter) may be synthesized. First the digitally-recorded speech is analyzed by programs which extract the attributes (parameters) of the speech for short time segments (.01 sec.). From these parameters the speech may be recreated in a form which resembles the original recording very closely. But for musical purposes the parameters are most often altered before synthesis. It is possible, for example, to change the natural pitch contour of the speech into a melodic line or to change the speech speed without altering the natural pitch level. The variety of musical patterns which may be created from the analytic parameters is limited only the composer's imagination.[9]

The dialogue itself takes place mostly between a man and a woman who discover a book that tells the story of their lives. They reflect on the seeming emptiness of their lives and in turn

---

[8] Ed M. Thieberger and Charles Dodge, "An Interview with Charles Dodge," *Computer Music Journal* 19, no. 1 (1995): 22.
[9] *Charles Dodge: Synthesized Voices*, CD liner notes (New York: New World Records, 2011).

become obsessed with the book. The book is personified with a markedly unnatural third voice that makes occasional interjections. As we will see, Dodge's method of speech analysis and synthesis allows for play between the speaking and singing voice, and between the human and the computer.

The basic conventions of recitative are "syllabic setting of the text and a close relationship between textual and musical phrases."[10] Recitative then serves as a sonic analog for prosodic dramatic delivery. This, to me, begs several questions with regard to Dodge's treatment of the line between recitative and speech: What do speech and recitative have in common? How do they differ? How do speech and music intersect and interact in recitative as a style? Dodge's recitative functions as a reflection on what separates the speaking voice from the singing voice. What musical features then does Dodge employ to construct recitative from speech synthesis?

On the general distinction between speech and singing, Richard Miller argues that

> because (1) breath management must be of a higher order in singing than in speaking, (2) the duration of the vowel is dissimilar in speaking and singing, (3) the compass of the singing voice exceeds that of speech inflection, (4) sung sound requires adjustments of breath energy to meet the shifting demands of pitch and intensity, and (5) the aesthetics of artistic singing require 'resonance balancing' beyond the needs of the speaking voice,…only in a limited sense—largely phonetic—does one sing '*come si parla*' [as one speaks].[11]

There is thus a difference in vocal delivery that limits what speech and singing can have in common. Regarding recitative specifically, in her survey of German 17th-century opera Judith Popovich Aikin reminds the reader that "in spite of the imitation of prose speech rhythms aspired to by the *parlando* style, opera recitative is, after all, a musical setting of verse in which rhyme

[10] Lawrence Michael Zbikowski, *Foundations of Musical Grammar* (New York, NY: Oxford University Press, 2017), 82.
[11] Richard Miller, *On the Art of Singing* (New York: Oxford University Press, 1996), 52.

patterns and rhythms, however flexible, are nevertheless a means of formal structuring."[12] This statement hints at the ultimate difference underlying speech and recitative: recitative is performative, an artistic representation of speech that is restricted by the resources provided by musical frameworks. Likewise, Pier Francesco Tosi, in his 1723 treatise *Opinioni de' cantori antichi e moderni* (which was translated and commented upon by Johann Friedrich Agricola), notes that theatrical recitative, in order to be beautiful, must be performed with the "stately decorum with which princes and those who consort with them speak."[13] This suggests a refined, heightened manner of performance beyond recreating everyday speech. Recitative is ultimately a *sung* gesture that adheres to musical conventions, which is what makes Dodge's recitative—wherein the voice began as speech—such a complex example.

For this discussion we can generalize about stylistic frameworks embedded in recitative. We can expect to hear rhythms that imitate speech—meaning that articulations in musical time are somewhat free-flowing and may not adhere to a clear meter. The form is likely to be through-composed, not relying on the kinds of direct or motivic repetitions we might expect from song. In terms of pitch, we will hear discrete musical pitches rather than the shifting vowel formants of speech, but the range of the voice might be limited to a relatively small range in order to imitate speech. Timbrally, we expect to hear the human voice at the focal point of the musical texture. As I will show, these expectations are at work in *The Story of Our Lives*, but Charles Dodge's individual compositional style tempers them, creating a unique interplay between speech and song.

---

[12] Judith Popovich Aikin, *A Language for German Opera: The Development for Forms and Formulas for Recitative and Aria in Seventeenth-Century German Libretti* (Wiesbaden: Harrassowitz in Kommission, 2002), 43.
[13] Johann Friedrich Agricola, *Introduction to the Art of Singing*, trans. Julianne Baird (Cambridge, England: Cambridge University Press, 1995), 171.

Dodge's means of composition, computer synthesis and analysis, brings up a further question: how does technology intervene in the relationship between speech and song in Dodge's recitative setting? Throughout *The Story of Our Lives*, we hear voices that play with the line between what is physically possible for the singing voice and what is a fantastical representation of the voice, unattainable by a live human performer. Not only does Dodge interrogate the line between speech and song in this work, but also the line between human and machine.

Much of Dodge's composition is like traditional recitative in its treatment of rhythm and pitch. The rhythms of the voice lie somewhere between the nonisochrony of speech and the proportional structure of musical divisions of time. Although I find tracking a clear musical meter to be difficult, at times the rhythmic delivery of the voice sounds as though it's been stretched into discernable musical rhythms. Take for example the opening line of the work from its first movement, "We Are Reading the Story of Our Lives" (0:03–0:17). In this excerpt, two synthesized voices—one male and one female—sing the phrase «We are reading the story of our lives, which takes place in a room.» The voices are spatially separated by hard panning, beginning with the woman's voice to the left and the man's voice to the right before switching places at 0:12. It is clear from listening that Dodge has made interventions into both the pitch and rhythmic content. Regarding pitch, the vowel formants have been altered so as to project more or less fixed pitches, and through this alteration Dodge treats the pitches musically, transforming the speaking voice into a singing voice. The two voices sing the same pitches an octave apart here. The delivery of the phrase sounds slower than typical speech. Notably, however, the onsets of syllables do not always align between the two voices.

Figure 5.1. Waveform analysis and rhythmic transcription, Charles Dodge, *The Story of Our Lives*, I. "We Are Reading the Story of Our Lives," 0:03–0:17

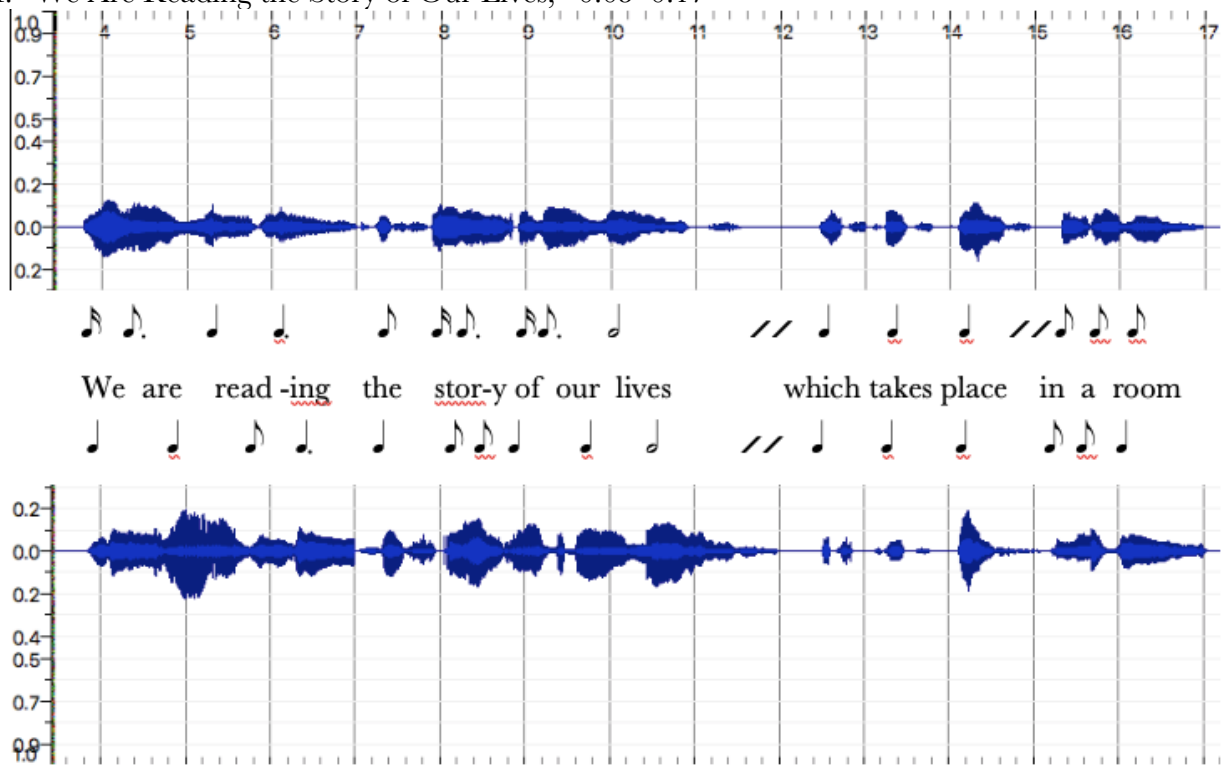
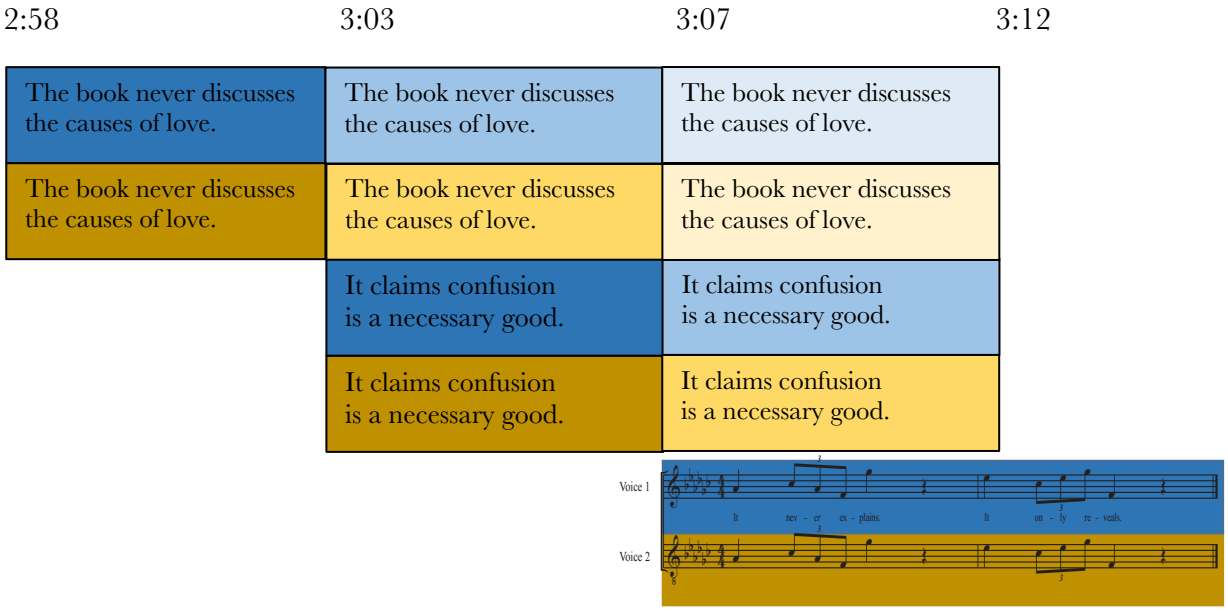


Figure 5.1 shows a waveform visualization with the left voice on top and the right voice on the bottom. Along with representations for each voice, I have provided a rhythmic transcription based on my hearing. While these rhythmicizations are somewhat rough—if we are to hear a tactus, it is quite flexible—they provide a visualization of the rhythmic nonalignment between voices and also demonstrate the potential for a musical hearing of rhythm in this phrase. More specifically, when the voices are in rhythmic unison at «which takes place» I hear evenly spaced, beat-like rhythms. The repeated sixteenth-dotted eighth rhythms in the left voice also help establish my hearing of a tactus. These building blocks of meter make me try to hear other rhythms in this proto-metrical context, which leads to my hearing of rubato. These rhythmic features ultimately align with my expectations of unmeasured, rhythmically free recitative. The use of two simultaneous voices, however, is unusual for the genre. Moments like these, which

subvert stylistic expectations, make *The Story of Our Lives* a particularly effective interrogation of the line between speech and song in recitative.

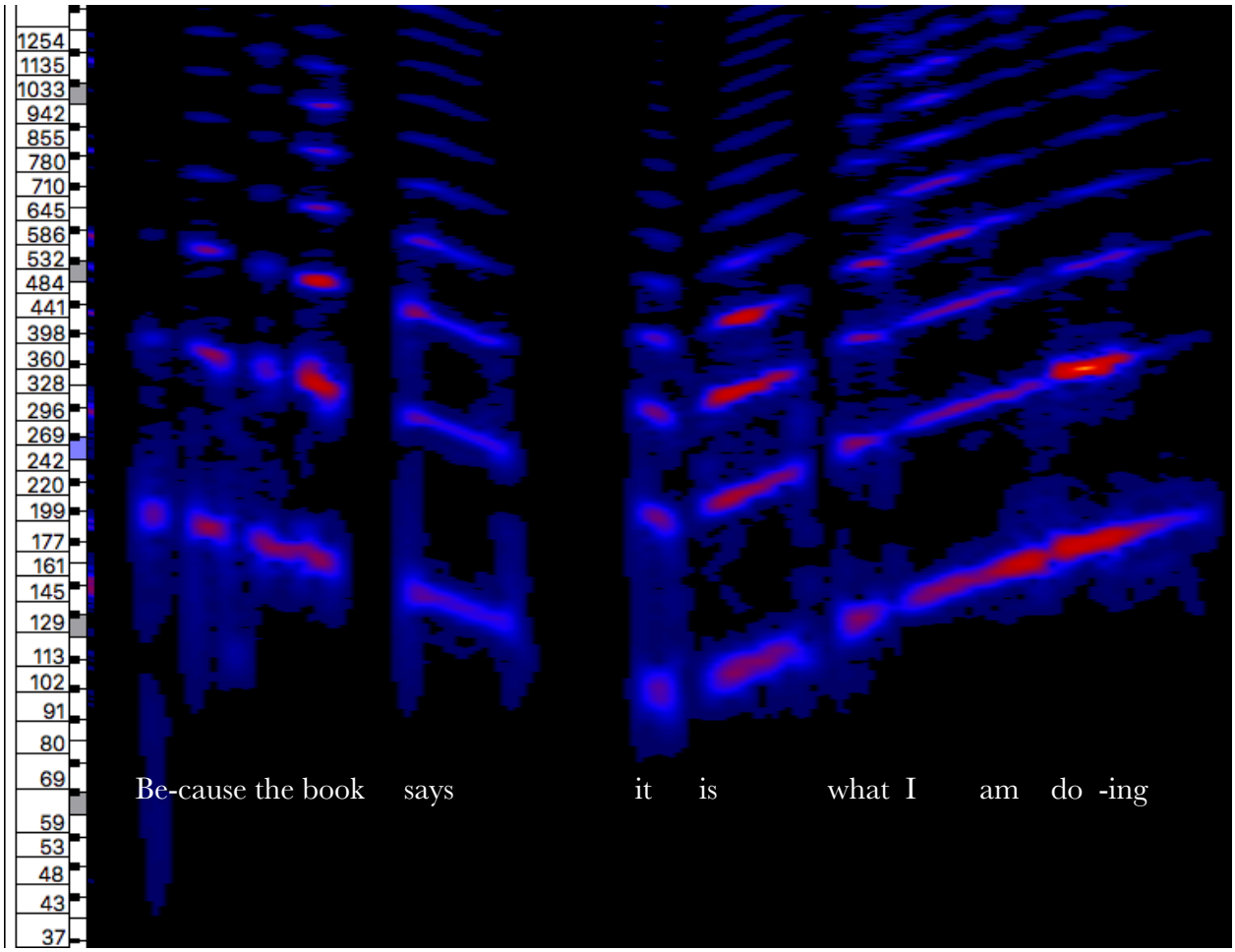
Figure 5.2. Transcription of polyphonic setting, Charles Dodge, *The Story of Our Lives*, V. "If Only There Were a Perfect Moment in the Book," 2:58–3:16



At times, Dodge removes parts of the stylistic framework of recitative, pushing the sound of the voice more toward speech but not entirely abandoning musical features. The end of the fifth movement, "If Only There Were a Perfect Moment in the Book" features, in the passage at 2:58–3:11, speech synthesis without the traditional musical elements of discrete pitch and proportional rhythms. Here the musical element is located in the texture: several voices (male and female) are overlaid in a sort of spoken counterpoint. Figure 5.2 shows a transcription of this section. (The darker colors represent greater loudness, and the spoken lines are represented with text while sung lines are notated.) The section begins with a single spoken line in both voices. As each new line enters, it is placed in the forefront of the texture by its volume and is then gets quieter on subsequent iterations as other voices are introduced. At 3:07, a doubled sung line akin

to the texture shown in Figure 5.1 returns, effectively cutting off the spoken polyphony at 3:10 and returning the piece to its normative musical state of two unaccompanied singing voices.

Figure 5.3. Spectrogram of «Because the book says it is what I am doing», Charles Dodge, *The Story of Our Lives*, II. "We Are Reading the Story of Our Lives," 0:25–0:29

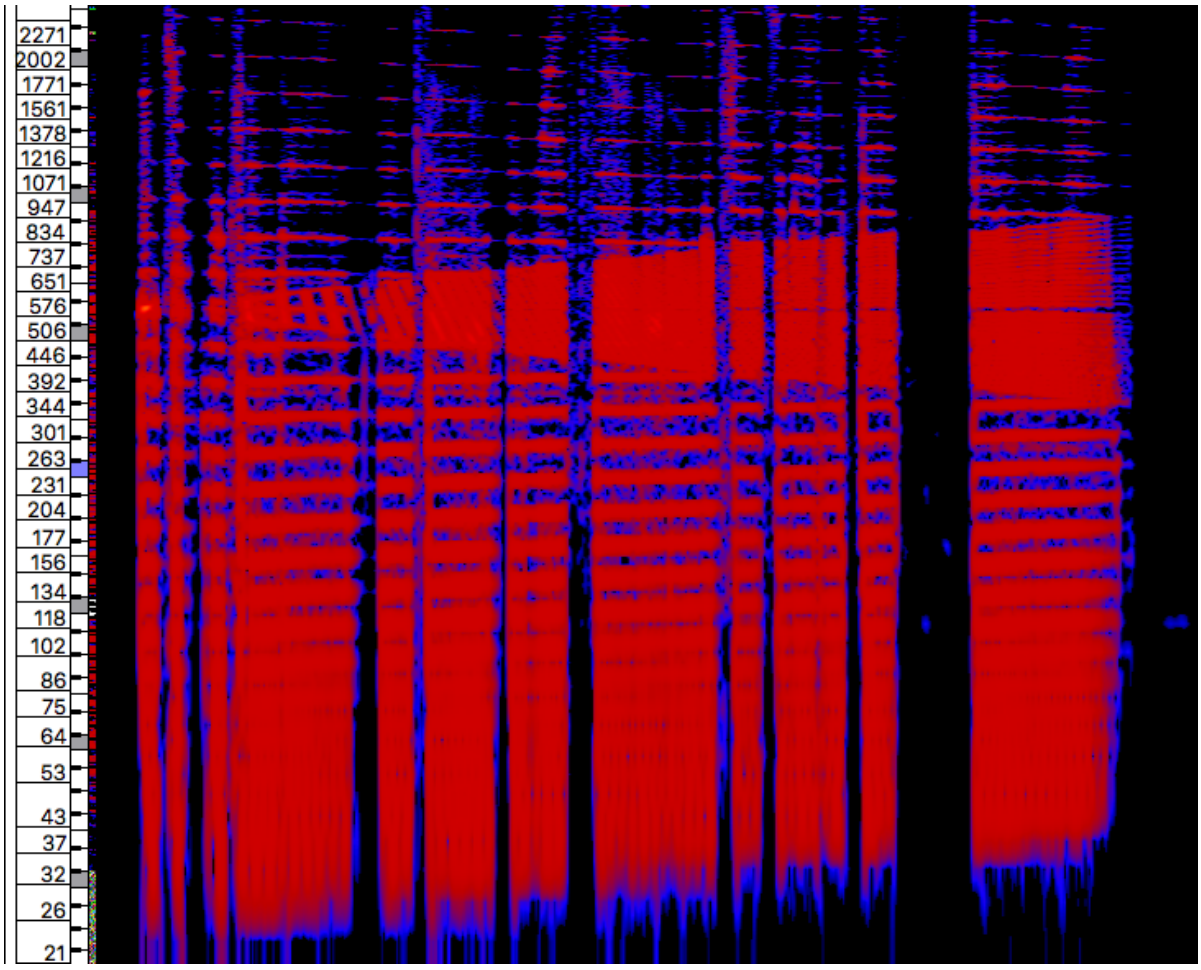


Furthermore, there are moments that call my attention to the synthesized nature of the voice and in doing so break the façade of song. There is, for instance, slippage from music to speech in the form of vowels whose pitches shift over their duration instead of holding discrete pitches. We can hear this at 0:25–0:29 of the second movement, also titled "We Are Reading the Story of Our Lives." In this example, there are two distinct phrases: «Because the book says» and

«it is what I am doing». These phrases are segmented by their contours, which are shown as spectrograms in Figure 5.3. The intonation of «Because the book says» falls over its duration, and that of «it is what I am doing» rises. In the spectrogram we can see that with these contours Dodge goes further than simply avoiding discrete musical pitches: he creates a linear passage through pitch space. Dodge has deployed pitch in a way that isn't natural in song *or* in speech. Here not only is the pretense of song broken, but indeed the whole pretense of an embodied voice is broken by this precise, mechanical movement of the synthesized voice through pitch space. The voice in *The Story of Our Lives* always sounds synthesized, but in instances like this it clearly pushes beyond the human.

Perhaps the most disruptive way Dodge reveals the technological mediation of the voice is through timbre, particularly in the third voice, the voice of the book itself. The book's voice is heavily processed, making it unnatural in a way that sounds insidious and otherworldly. The book's voice makes its first appearance in the second movement at 0:45–0:58, visualized in Figure 5.4 as a spectrogram. Compared with the voices of the man and woman, the timbre of the book's voice is incredibly thick and noisy. Consonants seems to pop and stand out in conspicuous ways. The voice has a whisper-like quality (brought about by the way Dodge's processing foregrounds sibilants), but with peculiar overtones that fill out the sound. We can hear simultaneously both rising low frequencies and falling high frequencies, making discerning pitch a disorienting experience. This eerie, inhuman voice highlights the lack of physical body for the book and underscores its dangerous nature as it makes the man and woman question their lives.

Figure 5.4. Spectrogram of the book's voice, «He put the pen down and turned a watched her reading the part about herself falling in love», Charles Dodge, *The Story of Our Lives*, II. "We Are Reading the Story of Our Lives," 0:45–0:58



As shown by these examples, Dodge uses an interplay between our stylistic expectations for recitative and his own technologically mediated take on the voice to create a hybrid on several fronts: between speech and song, between recitative and dramatic dialogue, between human and machine. His use of heterophonic rhythms for more or less "unison" passages plays on the room for expressive variation in the rhythms of speech. His deployment of plain speech in a polyphonic texture points to musical possibilities for speech beyond the formulaic sparse textures of recitative. His conspicuous, audible deployment of digital technology reminds us of

the new directions available for the voice in late 20th-century music, moving beyond what is bodily possible into a post-human, augmented realm of vocal possibility where both speech and song can exist simultaneously.

The Story of Our Lives points at the ways style can be leveraged between the levels of genre and personal compositional strategies. Dodge plays with the boundaries of what recitative can be. In doing so, he calls our attention to the communicative resources of music, particularly the ways in which they can diverge from or complement the communicative resources of speech.

Compositional homage as stylistic allusion in Peter Ablinger's "Arnold Schoenberg"

In this final section, I will consider musical works for recorded speech works that serve as homages to influential 20th-century composers whose voices are used as compositional objects. I'll focus on the "Arnold Schoenberg" movement from Peter Ablinger's *Voices and Piano*, which is based on a recorded voice memo of Schoenberg chastising his publisher. In this movement, Ablinger draws on Schoenberg's musical style as a means of interacting Schoenberg at multiple levels: as a composer, a historical figure, and, finally, as a human subject with a voice.

Ablinger's interest in building on the music of other composers is not a new concept. We saw in Chapter 3 that James Tenney employed Harry Partch's microtonal system and bespoke instruments in *Song 'n' Dance for Harry Partch* (1999). In a broader sense, we see in Western composition of the late 20th century a commitment to acknowledging the musical past through musical borrowing.[14] This is expressed through a variety of movements and techniques, and has been explored by a number of scholars. In her consideration of the 1960s collage compositions of

[14] Musical borrowing is of course not a new development. (Let's not forget the Renaissance paraphrase or parody mass.) For a broader historical overview of musical borrowing, see J. Peter Burkholder, "The Uses of Existing Music: Musical Borrowing as a Field," *Notes* 50, no. 3 (1994): 851–53.

Luciano Berio, George Rochberg, and Bernd Alois Zimmerman, Catherine Losada locates a link between the composers and their predecessors on two levels: in the directly quoted music of past composers and as "a response to and an outgrowth from the serial practices of the 1950s."[15] Yayoi Uno Everett explores György Ligeti's use of parody in *Le Grand Macabre* (1974–77), arguing that Ligeti employs references to earlier opera in way that creates new significations for these quotations in accordance with the narrative trajectory of the work.[16] Björn Heile sees Maurice Kagel's stylistic borrowings from various musical practices across the world in *Die Stücke der Windrose für Salonorchester* (1989) as an attempt at engagement with the issues of cross-cultural influence in an increasingly globalized world.[17] All of these practices fall into what Marina Lobanova dubs the "polystylistic situation" that began in Western music in the 1960s, a movement formed on the back of the collapse of the classical canon that is defined by an anti-hierarchical, open concept of style.[18] What we see in this polystylistic situation is a consciousness of the musical past intertwined with a desire to imbue new meanings into these sounds.

Frederic Jameson views the rise of pastiche and parody in art as a result of the disappearance of the individual subject and the unavailability of the personal style in the postmodern age.[19] But not all scholars see the use of quotation and borrowing in contemporary music in such bleak terms. Literature on sampling in particular often frames musical borrowing

---

[15] C. Catherine Losada, "Between Modernism and Postmodernism: Strands of Continuity in Collage Compositions by Rochberg, Berio, and Zimmermann," *Music Theory Spectrum* 31, no. 1 (2009): 97.
[16] Yayoi Uno Everett, "Signification of Parody and the Grotesque in György Ligeti's Le Grand Macabre," *Music Theory Spectrum* 31, no. 1 (2009): 26–56.
[17] Björn Heile, "'Transcending Quotation': Cross-Cultural Musical Representation in Mauricio Kagel's 'Die Stücke Der Windrose Für Salonorchester,'" *Music Analysis* 23, no. 1 (2004): 57–85.
[18] M. Lobanova, *Musical Style and Genre: History and Modernity* (Amsterdam: Harwood Academic, 2000), 157. See pages 153–169 for her full discussion of the polystylistic situation.
[19] Fredric Jameson, *Postmodernism, or, The Cultural Logic of Late Capitalism* (Durham: Duke University Press, 1991), 16.

as a historically aware act. Jesse Stewart cites DJ Spooky and his use of sampling as an example of Afro-postmodernism, a movement that "frequently uses pastiche, intertextuality, and irony as strategies of identity formation that remember and honor the cultural past, while at the same time working to construct visions of a better future."[20] DJ Spooky himself describes DJ culture and sampling as "a kind of archival impulse applied to a kind of hunter-gather milieu."[21]

Reflecting on postmodern musical borrowing in the late 20th century—and adopting a broad view of Western musical culture that includes both popular and concert music—Susan McClary argues that "far from being a time of failed art or the imitation of dead styles, it is a moment, I believe, of exuberant creativity, even when (perhaps especially when) its art refers most brazenly to earlier traditions."[22] According to Linda Hutcheon, "even the most self-conscious and parodic of contemporary works do not try to escape, but indeed foreground, the historical, social, and ideological contexts in which they have existed and continue to exist."[23] Along these lines, I see musical borrowing as an homage to influential musicians, along the lines of what McClary calls an "active negotiation with the cultural past for the sake of the here and now."[24] Thus works like Ablinger's "Arnold Schoenberg" can be seen as attempts the composer to recapture the voice of compositional predecessors not only through their words, but also through their compositional techniques.

---

[20] Jesse Stewart, "DJ Spooky and the Politics of Afro-Postmodernism," *Black Music Research Journal* 30, no. 2 (2010): 340f.
[21] Paul D. Miller, "In Through the Out Door: Sampling and the Creative Act," in *Sound Unbound: Sampling Digital Music and Culture*, ed. Paul D. Miller (Cambridge, Mass.: MIT Press, 2008), 13.
[22] Susan McClary, *Conventional Wisdom: The Content of Musical Form*, Bloch Lectures (Berkeley: University of California Press, 2000), 167.
[23] Linda Hutcheon, "The Politics of Postmodernism: Parody and History," *Cultural Critique*, no. 5 (1986): 183.
[24] Susan McClary, *Conventional Wisdom*, 168.

Along these lines, I argue that we should read Ablinger's "Arnold Schoenberg" (and many other works discussed in this dissertation) along the lines that DJ Spooky reads sampling writ-large: an "archival impulse" to engage with historical subjects in a way that foregrounds the subjects' embodied experience. Ablinger's treatment of Arnold Schoenberg's voice in this movement presents an example in which the composer has no personal connection to the speaker, but rather an awareness of the speaker's historical impact on Western concert music. Many movements in *Voices and Piano* use the voices of composers and musicians as a compositional object, and Ablinger's compositional strategies make reference to their unique musical personas. (See also "Hanns Eisler," "Morton Feldman," and "Billie Holiday.") Perhaps more importantly, Ablinger uses the piano accompaniment to refer to the speakers-as-composers. In these movements, Ablinger relies on stylistic allusion, borrowing not from any specific work but rather from a general style, to evoke the music of these subjects without directly quoting from their output.[25]

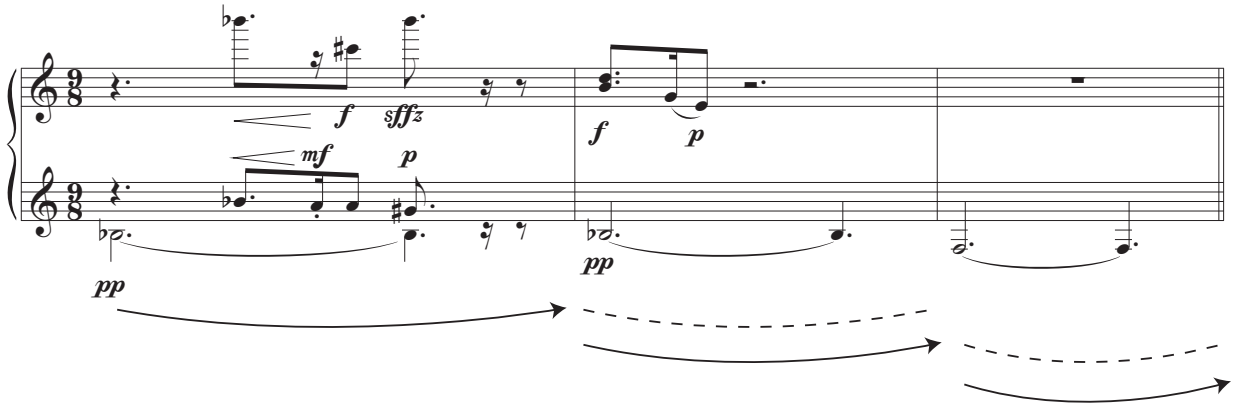
Recall from Chapter 3 that Ablinger uses measure-length pedal points like the ones seen in Figure 3.9 (reprinted here as Figure 5.5) to reinforce the notated meter and to add motivic repetition. While these pedal points are structurally crucial to Ablinger's piece, underscoring beginnings and endings of spoken phrases and promoting unity through motivic repetition, they also recall Schoenberg-as-composer by gesturing toward compositional strategies that Schoenberg used to shape the rhythmic structure of his instrumental music. Indeed, a number of authors have noted the importance of pedal points within Schoenberg's atonal output.[26] Pedal

[25] This follows J. Peter Burkholder's definition of stylistic allusion. See Burkholder, "The Uses of Existing Music," 854.
[26] Courtney S. Adams, "Techniques of Rhythmic Coherence in Schoenberg's Atonal Instrumental Works," *The Journal of Musicology* 11, no. 3 (1993): 337; J. Daniel Jenkins, "Schoenberg's Concept of 'Ruhende Bewegung,'" *Theory and Practice* 34 (2009): 93. See for

points are an example of *ruhende Bewegung*, static or resting motion, which Schoenberg defined as a "structuring element" of musical motion (along with accelerating and retarding).[27] Before the introduction of these pedal points in m. 10, the harmonic rhythm in this movement has been fairly quick—often just one eighth note in duration. The pedal points then serve to (1) build the hierarchy of the metrical framework and (2) still the harmonic motion, one of the uses of *ruhende Bewegung* described by J. Daniel Jenkins.[28]

Figure 5.5. The emergence of measure-level periodicity in mm. 10–12, Peter Ablinger, "Arnold Schoenberg," *Voices and Piano*, 0:18–0:23

5. "which I find terrible."



Texturally, Ablinger's musical moment shown in Figure 5.5 bears a notable similarity to mm. 1–4 of the minuet from Schoenberg's Suite for Piano, Op. 25 (shown in Figure 5.6). In the examples from both Ablinger's and Schoenberg's pieces, a pedal point underpins and provides contrast to the faster rhythmic motion found in the upper voices. Both involve unexpected wide leaps into the upper register as well as some simple stepwise motion in the primary melodic

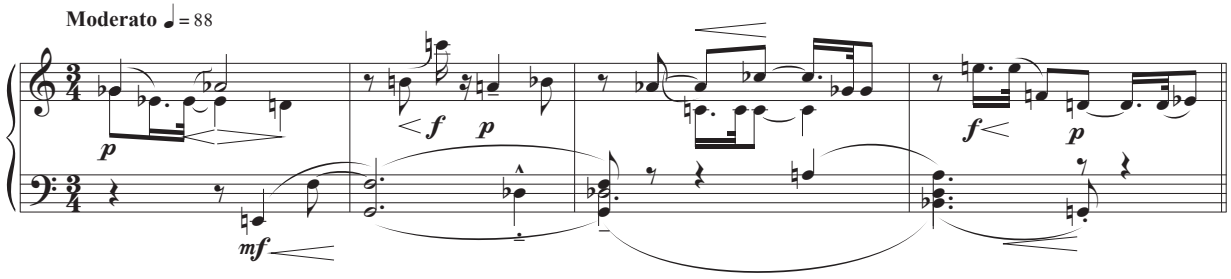
instance Op. 19, mvt. VI; mm. 1—12 of Op. 19, mvt. I; Op. 11, mvt. II; and m. 17 of Op. 11, mvt. III.

[27] Arnold Schoenberg, "[No Title]" (Vienna, Austria, n.d.), T53.06, Arnold Schönberg Center.

[28] J. Daniel Jenkins, "Issues of Form in Schoenberg's Atonal Period Vocal Music: Three Case Studies" (PhD dissertation, University of Rochester, Eastman School of Music, 2007), 172.

register. Each have a wide dynamic range with sudden contrasts between loud and soft. As should be evident, Ablinger does not quote Schoenberg directly, but he does quite effectively model Schoenberg's style through the way he organizes the rhythms, textures, and dynamics of his composition.

Figure 5.6. Arnold Schoenberg, Op. 25, mvt. V, mm. 1–4



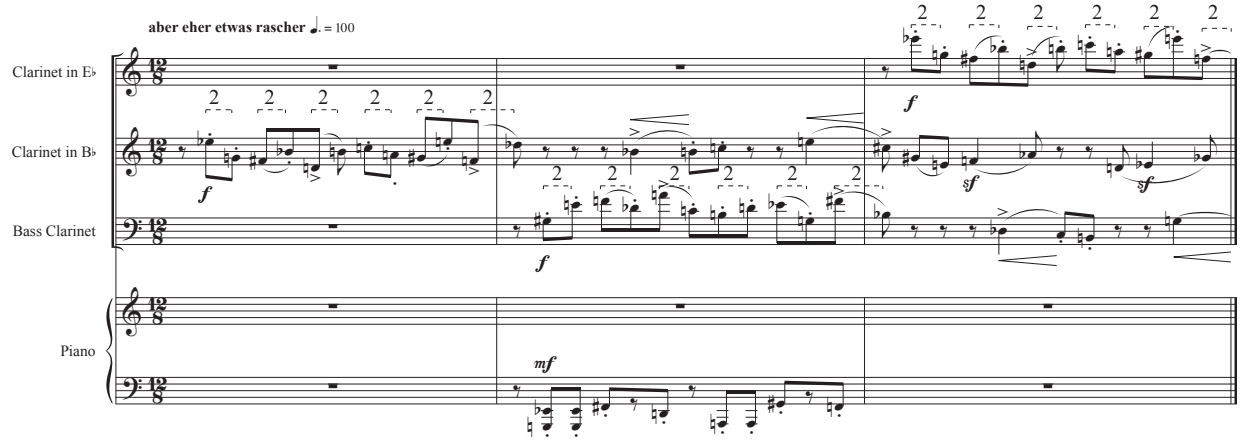
Another compositional technique used by Schoenberg in his atonal music—and one also exploited by Ablinger—is grouping dissonance at the level of subdivision. Ablinger instructs the performer to play the "Arnold Schoenberg" movement "like a gigue," and indeed Ablinger's rhythmic writing alludes to Schoenberg's rendering of the gigue topic.[29] Figure 5.7 offers two examples of grouping dissonance in Schoenberg's gigues. In the first (Figure 5.7a), which shows the opening measures of the fourth movement of the *Suite*, Op. 29 (1926), Schoenberg sets up duple groupings on the level of the eighth note. The result is a duple meter feel that persists until triple groupings are presented more clearly in the fourth measures. In the second (Figure 5.7b), which shows the opening of the fifth movement of the *Suite for String Orchestra* from 1934,

[29] Schoenberg's gigues include *Suite for String Orchestra* (1934), *Suite for Piano,* Op. 25 (1921–23), and *Suite*, Op. 29 (1926). The gigue as a dance topic carries with it certain rhythmic/metrical expectations. It should have a triple-meter feel, most often—but not always—written in 6/8. It often has dotted eighth notes on the beat (which, as noted previously, Ablinger uses as a rhythmic motive). Large melodic leaps and a fast harmonic rhythm are additional characteristics of the gigue. See Wye Jamison Allanbrook, *Rhythmic Gesture in Mozart: Le Nozze Di Figaro & Don Giovanni* (Chicago: University of Chicago Press, 1983), 41–43.

Schoenberg sets up grouping dissonances on two levels. In m. 1 the cello falls into duple eighth-note groupings, and switches to triple sixteenth-note groupings in the first two beats of m. 2. The first violin takes over the duple eighth-note groupings on the downbeat of m. 2, only to be passed the triple eighth-note groupings two beats later. Both of these excerpts illustrate that grouping dissonances at the level of subdivision are a common feature of Schoenberg's compositional style within the confines of the gigue topic. Ablinger employs this same stylistic feature in a number of passages in "Arnold Schoenberg." Perhaps the best example occurs in mm. 7–8, shown in Figure 5.8. This except contains both types of grouping dissonance identified in Figure 5.7, G3/2 at the level of the eighth note in the right hand of the first two beats of m. 7, and G2/3 at the level of the sixteenth note in the left hand of mm. 7–8. In this way Ablinger patterns his rhythmic structure after Schoenberg's rhythmic style within the gigue idiom, itself a parody of that classical dance topic.

Figure 5.7. Grouping dissonance in Schoenberg's gigues

a. Arnold Schoenberg, *Suite*, Op. 29, mvt. IV, mm. 1–3



b. Arnold Schoenberg, *Suite for String Orchestra*, mvt. V, mm. 1–2

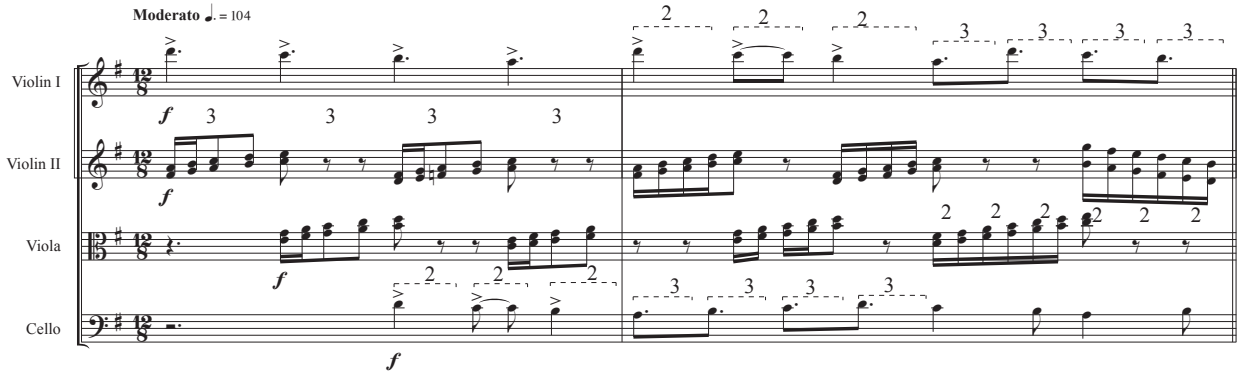
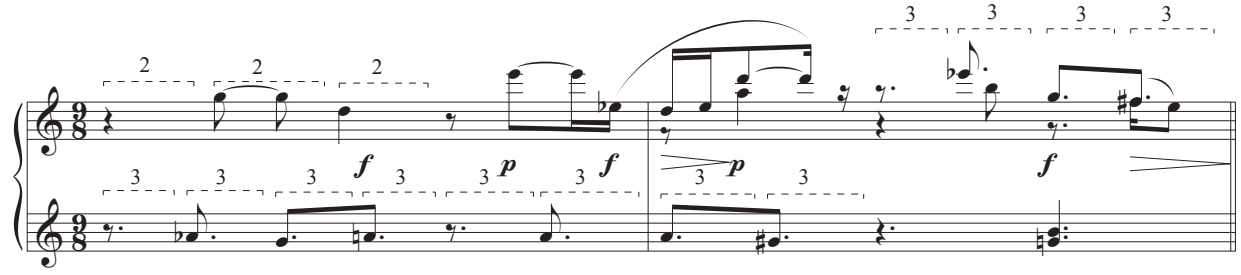


Figure 5.8. Grouping dissonance, Peter Ablinger, *Voices and Piano*, "Arnold Schoenberg," mm. 7–8



In these examples, we can see Ablinger's use of distinctive textural, registral, and rhythmic features from Schoenberg's compositional style as an act of parody. Schoenberg as a

composer is a looming authority figure for modern composition with whom composers in the Western concert tradition must grapple. His emancipation of dissonance is often touted as formative moment that shaped the trajectory of Western music in the 20th century.[30] By invoking Schoenberg's style, Ablinger draws a connection between himself and his speaker, highlighting Schoenberg's influence on the post-tonal milieu in which he is operating.

I also view Ablinger's treatment of Schoenberg's speaking voice as an allusion to Schoenberg's revolutionary relationship with the voice, particularly his development of Sprechstimme, an artistically heightened recitation style that lies between speech and song. Famously Schoenberg's choice to use Sprechstimme in *Pierrot Lunaire*, Op. 21 (1912) was influenced by German melodrama, which featured text spoken over musical accompaniment or musical interludes between spoken sections. According to Sharon Mabry, the conventions of melodrama appealed to Schoenberg because it "was quite successful as a vehicle for the interpretation of magical, supernatural, atmospheric, or mysterious texts, since the voice was not restricted to particular tones or rhythms and had total freedom of expression and nuance."[31] Ablinger's use of Schoenberg's voice mirrors this association with the supernatural, having Schoenberg perform new music from beyond the grave.[32] The fixity of the speech recording that Ablinger uses mirrors the control that Schoenberg maintained control over the performer's voice by explicitly notating pitches and rhythms in his Sprechstimme.

---

[30] For more on the history of Schoenberg's emancipation of dissonance and its historical implications, see Stephen Hinton, "The Emancipation of Dissonance: Schoenberg's Two Practices of Composition," *Music & Letters* 91, no. 4 (2010): 568–79.
[31] Sharon Mabry, *Exploring Twentieth-Century Vocal Music: A Practical Guide to Innovations in Performance and Repertoire* (Oxford: Oxford University Press, 2002), 78.
[32] It is interesting to note that audio recording was initially conceived of as a means of communicating with future generations. Jonathan Sterne sees this early use of recording technology as stemming from a wider cultural fascination with preservation and embalming in the Victorian era. See Jonathan Sterne, "The Resonant Tomb," in *The Audible Past: Cultural Origins of Sound Reproduction* (Durham: Duke University Press, 2003), 287–333.

Ultimately, Ablinger's piece holds many representations of Arnold Schoenberg. The musical setting showcases Schoenberg as a composer by borrowing from his compositional style, which in turn makes reference to his position in music history. The sort of mythical proportions that Schoenberg-as-composer has taken on, however, are undercut in Ablinger's piece by Schoenberg-as-speaker. The recording of Schoenberg's voice that Ablinger uses shows a very human view of Schoenberg—lashing out, muttering to himself (0:25–0:26), pausing mid-sentence to gather his thoughts. This is far from the polished composer persona that Schoenberg tended to project. The recorded voice, unedited by Ablinger, showcases Schoenberg the man. Ablinger's "Arnold Schoenberg" presents the listener with a multifaceted exploration of its subject, using many manifestations of the composer's persona to draw him posthumously into the collaboration.

How the composer's voice shapes the musicalization of speech

The composer's use of stylistic frameworks is a manifestation of the composer's voice, and yields a personalized way of treating, altering, and musicalizing the human speaking voice. Stylistic frameworks are a means of setting up for the musical expectations for the organization of sound. Remember that speech itself is a system of organizing sound with quite different rules than music. Stylistic frameworks serve then as a structure through which composers can adapt speech to fit within these frameworks (and also a lens through which the listener can sort the features of speech that are ambiguous in musical contexts into musical sound). I wish to focus on the means by which composers come to this adaptation of speech, considering how composers interact or collaborate with the recorded speaking voice in their compositional processes.

In Chapter 4, I proposed that composers from the late 20th century through today have included recorded voices in their musical works as a means of including more people in the

compositional process, albeit with complicated, uneven power dynamics. As this chapter has shown, it is not only the speaking voice that acts on the composer, but rather a whole network of influences. But while these actors have effect in constraining the musical possibilities, ultimately it is the composers themselves who have the most agency in these works. It is the composer's imagination that yields the final result—a piece of *music*. It is the composer who encourages the listener to hear speech in a different register, through a different system of organized sound.

These works transform speech (a basic, everyday form of communication) into art, a marked, enchanted sort of experience.[33] How then does the composer get us from speech to music? I propose that the composer's use and manipulation of the recorded voice begins an act of analysis, of hearing the speaking voice as potentially musical. By considering the potential of the speaking voice within her own milieu, a composer can imagine the possibility for imparting her musical hearing of speech onto other listeners.

Considering the composer as an analyst of the musical potential of speech adds another level of agency to the composer's role in the collaboration with the voice beyond just author. Seth Monahan's 2013 essay on agency in music analysis is useful for teasing out the types of agency involved in this kind of work. Monahan has proposed a four-tiered agential hierarchy for music analysis consisting of (from lowest to highest): *the individuated element* (themes, motives, etc.), *the work-persona* (the personification of the entire musical work), *the fictional composer*, and *the analyst*.[34] The analyst exists at the top of the hierarchy and exerts force on the other agent-classes. The work-persona and the individuated elements are objects that are given agency by the analyst.

---

[33] For more on art and technology as enchantment, see Alfred Gell, "The Technology of Enchantment and the Enchantment of Technology," in *Antropology, Art, and Aesthetics*, ed. Jeremy Coone and Anthony Shelton (Oxford, England: Oxford University Press, 1992), 40–63.
[34] Seth Monahan, "Action and Agency Revisited," *Journal of Music Theory* 57, no. 2 (2013): 327–47.

The fictional composer is the most complex hierarchical level: it is a person, one that has some control of the lower two levels, but it is ultimately constructed by the analyst rather than being a fully formed, autonomous human subject. As such, it has limited agency.

In the network of a musical work for recorded speech, there is likewise a hierarchy of agents at work, shown in Figure 5.9. Three of these present a one-to-one comparison. The composer serves as the analyst. Much like Monahan's analyst agent-class, an essential component of the composer's task is that of musical analysis. The composer must analyze the temporal, pitch, and timbral features of the speech recording and decide how to appropriately translate them into musical gestures. The recording itself functions as the work-persona, a contained expressive gesture, made into a unified entity by the boundedness of the recording. The individuated elements are moments (a spoken phrase, a word, a pitch contour, etc.) that the composer may tease out of the recording and put to use in the musical work.

Figure 5.9. Agency in the composer's musical analysis of speech

Monahan's agent-classes		Agent-classes in the musicalization of speech
1. Analyst	=	1. Composer-analyst
↓		↓ ↑
2. Fictional composer	=	2. Fictional composer, speaker, influences
↓		↓
3. Work-persona	=	3. Recording of speech
↓		↓
4. Individuated musical elements	=	4. Individuated elements of speech

The second level of the hierarchy contains several distinct agents. As this level deals with the representation of a person, it can be used to consider both the speaker, the composer's influences, the composer as she envisions herself as part of a sort of imagined collaborative act. Building on Monahan's work, George Adams sees Steve Reich's use of the recorded voice in *It's*

Gonna Rain as an intervention into traditional notions of the relationship between the composer, performance, work, and audience.[35] Adams argues that on the second level of the hierarchy, either the speaker/performer or composer may stand in for the fictional composer in Monahan's model. This is due to Reich's self-abnegation of his role as a composer in his process pieces, instead seeking to repress the subjectivity of the composer by crafting a musical process that essentially runs itself.[36] While other composers who work with speech as a compositional object have not been so vocal in their desire to displace their own subjectivity, I see the use of recorded speech as a means of relocating some level of authorial subjectivity onto another actor, the speaker. I add to this category other influences on the way a composer deploys musical frameworks: other musicians and composers, genres, and art movements. These influences, human or created by human society, exert some limited force on the composer in her decision-making process during composition.

In this situation, the composer-analyst is final arbiter of the soundworld of the musical work. The sonic features of the recording itself have influence over the possibilities for the musicalization of speech, but as we've seen in Chapter 3, there are technological means of overcoming these limitations created by the speaker. Influences similarly provide the basis for stylistic frameworks, but as we've seen with Charles Dodge and Peter Ablinger, the composer's unique musical strategies are based in part on building on and/or subverting stylistic expectations. Stylistic frameworks are the composer's means of grappling with the phonological features of the speaking voice and slotting them into musical systems of organization. Musical analysis along stylistic guidelines provides the basis for the act of collaboration with the speaking

---

[35] George Adams, "Listening to Conceptual Music: Technology, Culture, Analysis" (Chicago, University of Chicago, 2019), 191. See pages 182–212 for his complete discussion of agent-classes in Reich's musical process.
[36] Ibid., 198–207.

voice and positions the composer firmly at the top of this agential hierarchy as the actor who ultimately crafts the *music*.

Chapter 6: Conclusion

On April 11, 2020, the sketch comedy TV show *Saturday Night Live* aired its first of three "Saturday Night Live at Home" pre-recorded episodes that aired during the COVID-19 pandemic wherein cast members performed remotely. This first episode featured a skit, "Whatcha' Cookin' On?," in which current cast members Beck Bennett and Kyle Mooney, joined by musician and former *SNL* cast member Fred Armisen, use a (heavily edited) FaceTime video call to riff—both in the comedic and musical sense.[1] What begins as a spoken conversation gradually evolves over the course of the skit's 2 minutes and 15 seconds into a musical groove. As I will show, this skit demonstrates the ability of music in this repertoire of bring people together—those separated by distance, and in other cases, time.

The skit begins with a simple conversation between Bennett and Mooney as they try to come up with an idea for a comedy sketch. Around 0:34, Bennett and Mooney begin a rapid-fire back and forth of questions like «Whatcha' jammin' on?» and «Whatcha' workin' on?» The increase in the speed of the conversation, underscored by the corresponding onscreen cuts back and forth between the actors' faces, makes this feel less like an organic conversation than the first 33 seconds of the skit. This repetition of an idea, although not yet an exact repetition of a recorded segment, changes the tone of the communicative act and begins to lay the groundwork for a musical interpretation of the actors' voices. At 0:42–0:47, a musical phrase is created through alternating quick snippets of speech that emphasize the melodic quality and range of Bennett and Mooney's speech, which is spliced together in a way that sounds rhythmic. This same musical phrase is repeated exactly at 0:47–0:51, reinforcing this musical hearing. At 0:58 a

[1] Saturday Night Live, *Whatcha' Cookin' On* (New York: NBC Universal, 2020), https://www.youtube.com/watch?v=6zgSiNMSu8E&feature=youtu.be.

drum groove enters, bolstering the budding rhythmic framework, followed swiftly by bass and guitar, which provide a tonal pitch framework. (The images of dancing at 1:13–1:14 visually bolster this groove by suggesting a way in which the listener/viewer might interact with this music.) This groove continues until 1:17, when the musical accompaniment cuts out and the repeated snippets of speech are replaced by free-flowing conversation, returning again to unmarked speech. Finally, after adding Fred Armisen to the call, at 1:41 all of the previous musical markers return—the repeated speech snippets, the groove, the musical accompaniment—and the texture is thickened with additional speech snippets and instrumentation. At 2:02, the music once again dissipates for a final time into unmarked conversational speech as the actors decide to just use this conversation as their sketch for this week's episode.

I want to use this example as a lens through which to revisit the central questions that I laid out in Chapter 1: *How similar is music to speech? How does music adapt when confronted with spoken language as a compositional object? How can speech be heard as musical? How do music creators—composers, songwriters, producers, and musicians—use spoken language in the production of cultural products (musical works) that do not employ language as their primary communicative resource, and how do listeners apprehend this use? Can speech tip over into music only in selected instances, or is this kind of slippage a standard feature of human cognition?* Furthermore, both this comedy sketch and the larger phenomenon of using recorded speech structurally in music drives me to several other big-picture questions about contemporary music and music of the recent past, questions that I plan to explore more deeply in the future but for now will try to answer briefly in the context of "Whatcha' Cookin' On?": *What constitutes music? What are the bounds of the musical work? Who are the participants in the act of musicking?*

Between speech and music

First, let's address several closely related questions at once: *How similar is music to speech? How does music adapt when confronted with spoken language as a compositional object? How can speech be heard as musical?* As we've seen throughout this dissertation, music and speech are two wildly different ways of organizing sound; composers may, however, use strategies uniquely exploited by music—most crucially repetition—in order to emphasize the musical potential of speech rhythms and pitches.

In "Whatcha' Cookin' On?" music tips over gradually from speech to song, relying on a steady build-up of musical markers. This process begins a sort of "motivic" repetition at 0:34. It's not immediately apparent even when this repetition is introduced that music is the intended mode of delivery. The direct repetition beginning at 0:42 then places a new emphasis on rhythm and pitch. Finally, the introduction of musical instruments at 0:58 reinforces these suggested rhythmic and tonal frameworks. By this point there is no doubt that we are meant to hear this musically. Essentially, the listener's ability to hear speech musically relies on the presence of musical markers in the sound stream, and "Whatcha' Cookin' On?" demonstrates this by juxtaposing plain speech with music.

This leads to the broader question, *What constitutes music?* The simple answer is that music is organized sound, but so too is spoken language. The works discussed in this dissertation—particularly the way they use the human voice—highlight the range of communicative possibilities inherent in the voice: it is both a musical instrument and means of expressing natural language. Music ultimately relies on markers like repetition, discrete pitches, and metric hierarchy (although not all need be present) to signal itself as an artistic arrangement of sound.

Agents in the act of musicking

How do music creators—composers, songwriters, producers, and musicians—use spoken language in the production of cultural products (musical works) that do not employ language as their primary communicative resource, and how do listeners apprehend this use? This example begins with speech as the primary communicative resource. In the sound world of *SNL* speech is expected, music is not. As time goes on in "Whatcha' Cookin' On?" speech is waylaid—chopped up, repeated, and used in new sonic and communicative contexts—in service of music. Bennett, Mooney, and Armisen gradually lead the listener to hear their speech as musical, and just as quickly take away those markers and return to unmarked speech. What's clear, however, is that when the musical groove is initiated, the semantic meaning of speech is not nearly as important as our hearing of the musical structure. The musical moments of this skit rely on recordings of speech that we've already heard and interpreted in the context of conversation. Because of this, the listener can focus on the groove when it is initiated rather than try to understand the words being said, and yet can still understand the overall meaning behind the sketch. The music creators in this instance are able to guide the listener from speech to song, from semantic to phonological, and back.

While we're on the topic of music creators and music listeners, I turn to the question, *Who are the participants in the act of musicking?* In the music throughout this dissertation, we see a broad range of participants engaging with the final musical product—composers who notate music to be performed by others; musicians who perform these works and add their own layers of interpretation; composers who create a finished, fixed musical product; speakers who contribute to the structure of the music via the phonological structure of their utterances, which are then analyzed for its musical features by composers and listeners; listeners who hear both speech and music and understand their meanings in unique ways. We see in these works a sort of reciprocal

force between speakers and composers as authors of a musical work—the composers are in some ways beholden to the fixed, structural features of recorded speech but interpret them and add to them in ways that make an original artwork based in part on existing, nonmusical recordings; at the same time, the speakers leave an indelible mark on the final musical product, one that would not be the same with a different speaker or a different performance of the same words. Listeners can hear moments of recorded speech in these works as musical or not depending on the context. But, as we've seen throughout this dissertation, composers can use a general category of sound manipulation that pushes speech toward music in order to craft situations in which listeners are more likely to hear speech as musical sound.

The bounds of music

This idea leads us to the question, *Can speech tip over into music only in selected instances, or is this kind of slippage a standard feature of human cognition?* This question is the hardest to answer, and ultimately I think the answer lies in empirical cognitive research, beyond the scope of my research. I would like to posit, however, that the variety of examples of musicalized speech across genres and modes of delivery suggests that a wide range of people are inclined to hear the potential for musicality in the speaking voice. These people, musicians and composers, are then able to solidify these musical connections that they hear in speech by organizing and overlaying musical frameworks. The musical works that result from these endeavors then allow an even broader group of people, listeners, to hear the realized musical potential of speech recordings. This idea of slippage between speech and music leads me to my final big-picture question, *What are the bounds of the musical work?*

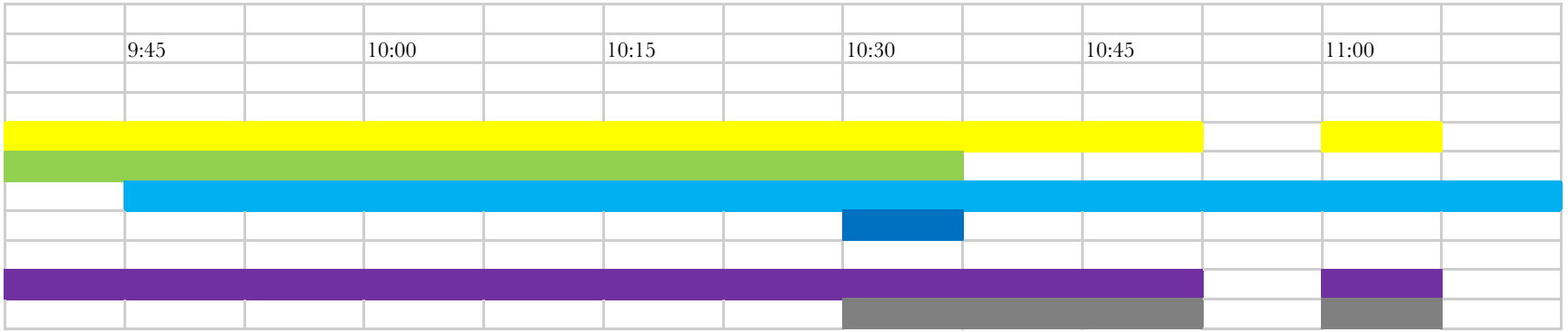
In "Whatcha' Cookin' On?" we hear moments of conversational speech, moments of music, and, particularly in 0:34–0:57, moments that fall in between. To some extent the

moments where speech slips over into music are crafted by creators Bennett, Mooney, and Armisen (and anyone on the *SNL* production team who may have edited the clips). But this slippage relies on the listener's ability to pick up on musical cues within speech and to process and interpret musical frameworks. This example illustrates the power that music creators have in curating a musical interpretation of speech that is legible as such to a listener.

Although for simplicity's sake I've opted to call the music discussed in this dissertation "musical works," "Whatcha' Cookin' On?" does raise questions about music and its uses. I'd hesitate to call the whole skit a "musical work," but *would* however use that term for Pamela Z's *Geekspeak*, which similarly plays between music and unmarked speech. In this dissertation, we see a wide range of uses for music: music intended for live performance, fixed tape works and album tracks, music for television, comedic videos. Each comes with a different intended audience and expected level of listener engagement. We see across this corpus the potential utility of music, both music for art's sake and music for other purposes. I believe that because of this music's liminal space between spoken communication and musical art, this compositional technique lends itself to a variety of uses. Ultimately, I see its appeal in its ability to bring various actors together over distances of time and space, to enable a distinctly modern means of displaced, virtual interpersonal interaction.

Appendix 4.1. Form diagram, Jacob TV, "Billie"


Development
4:45
5:00
5:15
5:30
5:45
6:00
6:15
"Doubles"
Accompanies
Singing
Singing
6:30
6:45
7:00
7:15
7:30
7:45
A'
8:00
8:15
8:30
8:45
9:00
9:15
9:30
Speaking


9:45
10:00
10:15
10:30
10:45
11:00

Bibliography

Abbate, Carolyn. *Unsung Voices: Opera and Musical Narrative in the Nineteenth Century*. Princeton, N.J.: Princeton University Press, 1991.

Ablinger, Peter. "Peter Ablinger - Quadraturen," July 28, 2016. http://ablinger.mur.at/docu11.html.

Ablinger, Peter, and Aljoscha Hofmann. "A Letter From Schoenberg." Peter Ablinger, September 31, 2002. http://ablinger.mur.at/txt_qu3schoenberg.html.

Adams, Courtney S. "Techniques of Rhythmic Coherence in Schoenberg's Atonal Instrumental Works." *The Journal of Musicology* 11, no. 3 (1993): 330–56.

Adams, George. "Listening to Conceptual Music: Technology, Culture, Analysis." University of Chicago, 2019.

Adult Swim. "Ants In My Eyes Johnson | Rick and Morty | Adult Swim." YouTube, July 22, 2015. https://www.youtube.com/watch?v=G4BkGJj5BXg.

Agricola, Johann Friedrich. *Introduction to the Art of Singing*. Translated by Julianne Baird. Cambridge, England: Cambridge University Press, 1995.

Aikin, Judith Popovich. *A Language for German Opera: The Development for Forms and Formulas for Recitative and Aria in Seventeenth-Century German Libretti*. Wiesbaden: Harrassowitz in Kommission, 2002.

Allanbrook, Wye Jamison. *Rhythmic Gesture in Mozart: Le Nozze Di Figaro & Don Giovanni*. Chicago: University of Chicago Press, 1983.

Anderson, Perry. *The Origins of Postmodernity*. London: Verso, 1998.

Ansell, Derek. "Billie Holiday: Lady In Satin | Jazz Journal." *Jazz Journal* (blog), September 17, 2019. https://jazzjournal.co.uk/2019/09/17/billie-holiday-lady-in-satin-2/.

Barrett, G. Douglas. *After Sound: Toward a Critical Music*. New York: Bloomsbury Academic, an imprint of Bloomsbury Publishing Inc., 2016.

Bartley, Wendalyn. "Concert Report: The Fine Art of Social Consciousness – 'James Tenney: Resistance.'" The WholeNote, October 24, 2017. https://www.thewholenote.com/index.php/other-media/blog/concert-reviews/27382-concert-report-the-fine-art-of-social-consciousness-james-tenney-resistance.

Beirens, Maarten. "Voices, Violence and Meaning: Transformations of Speech Samples in Works by David Byrne, Brian Eno and Steve Reich." *Contemporary Music Review* 33, no. 2 (2014): 210–222.

Billie Holiday Interviewed on Peacock Alley TV Show (1956). Accessed September 28, 2019. https://www.youtube.com/watch?v=JDl1fkwsDXM.

JacobTV.net. "Bio." Accessed September 26, 2019. https://www.jacobtv.net/composer.html.

Burkholder, J. Peter. "The Uses of Existing Music: Musical Borrowing as a Field." *Notes* 50, no. 3 (1994): 851–70.

Butler, Mark J. *Unlocking the Groove: Rhythm, Meter, and Musical Design in Electronic Dance Music*. Bloomington: Indiana University Press, 2006.

Campion, Thomas. "Observations in the Art of English Poesie [1602]." In *Campion's Works*, edited by Sylvanus Percival Vivian, 31–56. Oxford: Clarendon Press, 1909.

Carr, Philip. *A Glossary of Phonology*. Glossaries in Linguistics. Edinburgh: Edinburgh University Press, 2008.

Charles Dodge: Synthesized Voices. CD. New York: New World Records, 2011.

Chow, Ivan, and Steven Brown. "A Musical Approach to Speech Melody." *Frontiers in Psychology* 9 (March 2018).

Christiansen, Paul. "The Meaning of Speech Melody for Leos Janáček." *Journal of Musicological Research* 23 (2003): 241–63.

Clarke, Donald. *Billie Holiday: Wishing on the Moon*. Cambridge, MA: Da Capo Press, 2009.

Clarke, Eric F., and Mark Doffman. *Distributed Creativity: Collaboration and Improvisation in Contemporary Music*. New York, NY: Oxford University Press, 2017.

Clarke, Michael. "Jonathan Harvey's Mortuos Plango, Vivos Voco." In *Analytical Methods of Electroacoustic Music*, edited by Mary Simoni, 111–143. New York: Routledge, 2006.

Code, David Loberg. "Observations in the Art of Speech: Paul Lansky's Six Fantasies." *Perspectives of New Music* 28, no. 1 (1990): 144–169.

Condillac, Etienne Bonnot de. *An Essay on the Origin of Human Knowledge; Being a Supplement to Mr. Locke's Essay on the Human Understanding.* Translated by Thomas Nugent. History of Psychology Series. Gainesville, Fla.: Scholars' Facsimiles & Reprints, 1971.

Cone, Edward T. *The Composer's Voice*. Berkeley: University of California Press, 1974.

Connor, Steven. "The Decomposing Voice of Postmodern Music." *New Literary History* 32, no. 3 (2001): 467–483.

Cook, Perry R. "Singing Voice Synthesis: History, Current Work, and Future Directions." *Computer Music Journal* 20, no. 3 (1996): 38–46.

Cutler, Anne, Delphin Dahan, and Wilma Van Donselaar. "Prosody in the Comprehension of Spoken Language: A Literature Review." *Language and Speech* 40, no. 2 (1997): 141–201.

Dall'Ara-Majek, Ana. "Le Trésor de La Langue de René Lussier: Documentation d'un Voyage, de La Quête Linguistique Au Débat Politique." *Circuit: Musiques Contemporaines* 28, no. 3 (2018): 71–86.

Davies, Stephen. *Musical Meaning and Expression*. Ithaca: Cornell University Press, 1994.

Deutsch, Diana. *Phantom Words and Other Curiosities*. CD. Vol. 002. Philomel Records, 2003.

Deutsch, Diana, Trevor Henthorn, and Rachael Lapidis. "Illusory Transformation from Speech to Song." *The Journal of the Acoustical Society of America* 129, no. 4 (April 2011): 2245–52.

Dibben, Nicola. *Björk*. Icons of Pop Music. Bloomington: Indiana University Press, 2009.

Sub Rosa. "Dj Spooky (That Subliminal Kid)." Accessed December 12, 2019. https://www.subrosa.net/en/catalogue/soundworks/dj-spooky-that-subliminal-kid_01.html.

Dockery, David. "Ants in My Eyes Johnson w/Drums." YouTube, July 26, 2017. https://www.youtube.com/watch?v=G8UT6nkSWT4.

Dudley, Homer. "The Automatic Synthesis of Speech." *Proceedings of the National Academy of Sciences of the United States of America* 25, no. 7 (1939): 377–383.

Early Works. New York, N.Y.: Nonesuch, 1987.

Eaton, Rebecca M. Doran. "Marking Minimalism: Minimal Music as a Sign of Machines and Mathematics in Multimedia." *Music and the Moving Image* 7, no. 1 (2014): 3–23.

Eidsheim, Nina Sun. *The Race of Sound: Listening, Timbre, and Vocality in African American Music*. Refiguring American Music. Durham: Duke University Press, 2019.

Everett, Yayoi Uno. "Signification of Parody and the Grotesque in György Ligeti's Le Grand Macabre." *Music Theory Spectrum* 31, no. 1 (2009): 26–56.

Fairbanks, Grant, and Patti Grubb. "A Psychophysical Investigation of Vowel Formants." *Journal of Speech and Hearing Research* 4, no. 3 (September 1961): 203–19.

Falk, Simone, Tamara Rathcke, and Simone Dalla Bella. "When Speech Sounds Like Music." *Journal of Experimental Psychology: Human Perception and Performance* 40, no. 4 (August 2014): 1491–1506.

Feldman, Martha. "The Interstitial Voice: An Opening." *Journal of the American Musicological Society* 68, no. 3 (2015): 653–59.

Figueras, Marc Anthony. "'Ants In My Eyes Johnson' with Guitar." YouTube, August 11, 2017. https://www.youtube.com/watch?v=QBgjsm0XQ4Y.

Frigo, Connie. *Billie*. CD. Shining City. Aalsmeer, The Netherlands: Basta, 2007.

Garner, Tom A., and Mark Grimshaw. "Sonic Virtuality: Understanding Audio in a Virtual World." In *The Oxford Handbook of Virtuality*, edited by Mark Grimshaw, 364–377. Oxford: Oxford University Press, 2014.

Gell, Alfred. "The Technology of Enchantment and the Enchantment of Technology." In *Antropology, Art, and Aesthetics*, edited by Jeremy Coone and Anthony Shelton, 40–63. Oxford, England: Oxford University Press, 1992.

Gjerdingen, Robert O., and David Perrott. "Scanning the Dial: The Rapid Recognition of Music Genres." *Collected Work: Journal of New Music Research. XXXVII/2 (June 2008): From Genres to Tags: Music Information Retrieval in the Age of Social Tagging. Published by: Routledge, Abingdon, Oxfordshire, United Kingdom, 2008. (AN: 2008-07425).* 37, no. 2 (2008): 93–100.

Gopinath, Sumanth. "The Problem of the Political in Steve Reich's Come Out." In *Sound Commitments: Avant-Garde Music and the Sixties*, edited by Robert Adlington, 121–144. Oxford: Oxford University Press, 2009.

Grisey, Gérard, and Joshua Fineberg. "Did You Say Spectral?" *Contemporary Music Review* 19, no. 3 (January 2000): 1–3.

Gussmann, Edmund. *Phonology: Analysis and Theory*. Cambridge, U.K.: Cambridge University Press, 2002.

Hanninen, Dora A. *A Theory of Music Analysis: On Segmentation and Associative Organization*. Rochester, NY: University of Rochester Press, 2012.

Harvey, Jonathan. "Mortuos Plago, Vivos Voco: A Realization at IRCAM." *Computer Music Journal* 5, no. 4 (1981): 22–24.

Hasegawa, Robert. "Gérard Grisey and the 'Nature' of Harmony." *Music Analysis* 28, no. 2–3 (2009): 349–371.

Hasty, Christopher Francis. *Meter as Rhythm*. New York: Oxford University Press, 1997.

Hayles, N. Katherine. "Voices out of Bodies, Bodies out of Voices: Audiotape and the Production of Subjectivity." In *Sound States: Innovative Poetics and Acoustical Technologies*, edited by Adalaide Morris, 74–96. Chapel Hill: University of North Carolina Press, 1997.

Heile, Björn. "'Transcending Quotation': Cross-Cultural Musical Representation in Mauricio Kagel's 'Die Stücke Der Windrose Für Salonorchester.'" *Music Analysis* 23, no. 1 (2004): 57–85.

Helmholtz, Hermann von. *On the Sensations of Tone as a Physiological Basis for the Theory of Music.* Translated by Alexander John Ellis. New York: Dover Publications, 1863.

Hinton, Stephen. "The Emancipation of Dissonance: Schoenberg's Two Practices of Composition." *Music & Letters* 91, no. 4 (2010): 568–79.

Hirz, Rozalie. "Frequency-Based Compositional Techniques in the Music of Tristan Murail." In *Contemporary Compositional Techniques and OpenMusic*, 93–196. Paris: IRCAM, 2009.

Hodges, Nicolas. *Voices and piano*. CD. Vienna, Austria: Kairos Music, 2009.

Holiday, Billie, and Ray Ellis and His Orchestra. *Lady in Satin*. CD. New York: Columbia, 1997.

Horlacher, Gretchen. "Multiple Meters and Metrical Processes in the Music of Steve Reich." *Intégral* 14/15 (2000/2001): 265–97.

Howard, Luke. "The Voice(-Over) of God: Some Thoughts on the Disembodied Voice in Contemporary Music." *The Open Space Magazine* 1 (1999): 109–116.

Huang, Hao, and Rachel Huang. "She Sang as She Spoke: Billie Holiday and Aspects of Speech Intonation and Diction." *Jazz Perspectives* 7, no. 3 (December 2013): 287–302.

Huron, David. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, Mass.: MIT Press, 2006.

Hutcheon, Linda. "The Politics of Postmodernism: Parody and History." *Cultural Critique*, no. 5 (1986): 179–207.

Ilie, Gabriella, and William Forde Thompson. "A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect." *Music Perception: An Interdisciplinary Journal* 23, no. 4 (2006): 319–30.

Iverson, Jennifer. *Electronic Inspirations: Technologies of the Cold War Musical Avant-Garde*. The New Cultural History of Music. New York, NY: Oxford University Press, 2019.

———. "Invisible Collaboration: The Dawn and Evolution of Elektronische Musik." *Music Theory Spectrum* 39, no. 2 (2017): 200–222.

Jaisin, Kankamol, Rapeepong Suphanchaimat, Mauricio A. Figueroa Candia, and Jason D. Warren. "The Speech-to-Song Illusion Is Reduced in Speakers of Tonal (vs. Non-Tonal) Languages." *Frontiers in Psychology* 7 (2016).

Jameson, Fredric. *Postmodernism, or, The Cultural Logic of Late Capitalism*. Durham: Duke University Press, 1991.

Jenkins, J. Daniel. "Schoenberg's Concept of 'Ruhende Bewegung.'" *Theory and Practice* 34 (2009): 87–105.

———. "Issues of Form in Schoenberg's Atonal Period Vocal Music: Three Case Studies." Ph.D., University of Rochester, Eastman School of Music, 2007.

Juslin, Patrik N., and Petri Laukka. "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?" *Psychological Bulletin* 129, no. 5 (2003): 770–814.

King, Jeffrey C., and Jason Stanley. "Semantics, Pragmatics, and the Role of Semantic Content." In *Semantics versus Pragmatics*, edited by Zoltan Gendler Szabo, 111–161. Oxford: Oxford University Press, 2005.

Klein, Michael Leslie. *Music and the Crises of the Modern Subject*. Musical Meaning and Interpretation. Bloomington, Indiana: Indiana University Press, 2015.

Knoop, Mark. *Popular Contexts & Voices and Piano*. CD. Vol. SR382. Brussels: Sub Rosa, 2013.

Kramer, Lawrence. *Expression and Truth: On the Music of Knowledge*. Berkeley: University of California Press, 2012.

———. *Musical Meaning: Toward a Critical History*. Berkeley: University of California Press, 2002.

Krebs, Harald. *Fantasy Pieces: Metrical Dissonance in the Music of Robert Schumann*. New York: Oxford University Press, 1999.

Kreiman, Jody, and Diana Sidtis. *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Malden, MA: Wiley-Blackwell, 2011.

Kronos Quartet, and Pat Metheny. *Different Trains / Electric Counterpoint*. CD. 79176-2. New York: Elektra Nonesuch, 1989.

Krumhansl, Carol L. "Plink: 'Thin Slices' of Music." *Music Perception: An Interdisciplinary Journal* 27, no. 5 (2010): 337–54.

Kurland, Daniel. "'Rick and Morty' Threw Out the Script to Make One of Their Best Episodes." Vulture, May 24, 2016. http://www.vulture.com/2016/05/rick-and-morty-threw-out-the-script-to-make-one-of-their-best-episodes.html.

Lacasse, Serge. "The Phonographic Voice: Paralinguistic Features and Phonographic Staging in Popular Music Singing." In *Recorded Music: Performance, Culture and Technology*, edited by Amanda Bayley, 225–251. Cambridge, UK: Cambridge University Press, 2010.

Lansky, Paul. *Fantasies and Tableaux*. CD. CRI 683. New York: Composers Recordings Inc., 1994.

LaRue, Jan. *Guidelines for Style Analysis*. 2nd ed. Warren, Mich.: Harmonie Park Press, 1992.

Lerdahl, Fred, and Ray S. Jackendoff. *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press, 1983.

Leydon, Rebecca. "Towards a Typology of Minimalist Tropes." *Music Theory Online* 8, no. 4 (2002). http://www.mtosmt.org/issues/mto.02.8.4/mto.02.8.4.leydon.html.

Lidov, David. *Is Language a Music?: Writings on Musical Form and Signification*. Bloomington: Indiana University Press, 2005.

Lieberman, Philip, William Katz, Allard Jongman, Roger Zimmerman, and Mark Miller. "Measures of the Sentence Intonation of Read and Spontaneous Speech in American English." *The Journal of the Acoustical Society of America* 77, no. 2 (February 1985): 649–57.

List, George. "The Boundaries of Speech and Song." *Ethnomusicology* 7, no. 1 (1963): 1–16.

Lobanova, M. *Musical Style and Genre: History and Modernity*. Amsterdam: Harwood Academic, 2000.

London, Justin. "Cognitive and Aesthetic Aspects of Metrical Ambiguity." Presented as a colloquium talk, University of Pennsylvania, Fall 2008.

———. *Hearing in Time: Psychological Aspects of Musical Meter*. New York: Oxford University Press, 2004.

Losada, C. Catherine. "Between Modernism and Postmodernism: Strands of Continuity in Collage Compositions by Rochberg, Berio, and Zimmermann." *Music Theory Spectrum* 31, no. 1 (2009): 57–100.

Lussier, René. *Le trésor de la langue*. Montréal: Ambiances Magnétiques, 1989.

Mabry, Sharon. *Exploring Twentieth-Century Vocal Music: A Practical Guide to Innovations in Performance and Repertoire*. Oxford: Oxford University Press, 2002.

Madrid, Alejandro L. *In Search of Julián Carrillo and Sonido 13*. Currents in Latin American & Iberian Music. New York, NY: Oxford University Press, 2015.

Manabe, Noriko. "We Gon' Be Alright? The Ambiguities of Kendrick Lamar's Protest Anthem." *Music Theory Online* 25, no. 1 (May 2019). http://mtosmt.org/issues/mto.19.25.1/mto.19.25.1.manabe.html.

Manuel, Peter. "Music as Symbol, Music as Simulacrum: Postmodern, Pre-Modern, and Modern Aesthetics in Subcultural Popular Musics." *Popular Music* 14, no. 2 (1995): 227–39.

Margulis, Elizabeth H., Rhimmon Simchy-Gross, and Justin L. Black. "Pronunciation Difficulty, Temporal Regularity, and the Speech-to-Song Illusion." *Frontiers in Psychology* 6 (2015).

Margulis, Elizabeth Hellmuth. *On Repeat: How Music Plays the Mind*. New York, NY: Oxford University Press, 2014.

Maus, Fred Everett. "Agency in Instrumental Music and Song." *The College Music Symposium* 29 (1989): 31–43.

McClary, Susan. *Conventional Wisdom: The Content of Musical Form*. Berkeley: University of California Press, 2000.

McFarland, Kevin. "The Secret Behind Kimmy Schmidt's Perfect Theme Song." *Wired*, March 12, 2015. https://www.wired.com/2015/03/kimmy-schmidt-theme-song/.

McNeill, David. *Gesture and Thought*. Chicago: University of Chicago Press, 2005.

Merrill, Julia, and Pauline Larrouy-Maestri. "Vocal Features of Song and Speech: Insights from Schoenberg's Pierrot Lunaire." *Frontiers in Psychology* 8 (2017): 1108.

Mertens, Wim. *American Minimal Music: La Monte Young, Terry Riley, Steve Reich, Philip Glass*. Translated by J. Hautekiet. London: Kahn & Averill, 1983.

Meyer, Leonard B. *Style and Music: Theory, History, and Ideology*. Philadelphia: University of Pennsylvania Press, 1989.

Midgette, Anne. "Dutch Composer Samples Pop Culture and Gives It a Melody." *The New York Times*, May 4, 2007, sec. Music. https://www.nytimes.com/2007/05/04/arts/music/04jaco.html.

Miller, Paul D. (DJ Spooky), ed. *Sound Unbound: Sampling Digital Music and Culture*. Cambridge, Mass.: MIT Press, 2008.

Miller, Paul D. (DJ Spooky). "In Through the Out Door: Sampling and the Creative Act." In *Sound Unbound: Sampling Digital Music and Culture*, edited by Paul D. (DJ Spooky) Miller, 5–19. Cambridge, Mass.: MIT Press, 2008.

Miller, Richard. *On the Art of Singing*. New York: Oxford University Press, 1996.

Mills, Mara. "Media and Prosthesis: The Vocoder, the Artificial Larynx, and the History of Signal Processing." *Qui Parle: Critical Humanities and Social Sciences* 21, no. 1 (2012): 107–49.

Monahan, Seth. "Action and Agency Revisited." *Journal of Music Theory* 57, no. 2 (2013): 321–71.

Monelle, Raymond. *Linguistics and Semiotics in Music*. Chur, Switzerland: Harwood Academic, 1992.

Murail, Tristan. "The Revolution of Complex Sounds." *Contemporary Music Review* 24, no. 2–3 (April 2005): 121–35.

Murphy, Nancy. "The Times Are A-Changin': Metric Flexibility and Text Expression in 1960s and 1970s Singer-Songwriter Music." *Music Theory Spectrum*, forthcoming.

Naiman, Tiffany. "Singing at Death's Door: Late Style, Disability, and the Temporality of Illness in Popular Music." Ph.D., University of California, Los Angeles, 2017.

Netflix. "Unbreakable Kimmy Schmidt | Bankston Interview | Netflix." YouTube, March 9, 2015. https://www.youtube.com/watch?time_continue=30&v=HUk6GmqfqnY.

O'Callaghan, James. "Mimetic Instrumental Resynthesis." *Organised Sound* 20, no. 2 (2015): 231–40.

———. "Spectral Music and the Appeal to Nature." *Twentieth-Century Music* 15, no. 1 (February 2018): 57–73.

Ogden, Richard. *An Introduction to English Phonetics*. Second Edition. Edinburgh Textbooks on the English Language. Edinburgh: Edinburgh University Press, 2017.

Ondishko, Denise Michelle. "Six Fantasies on a Poem by Thomas Campion: Synthesis and Evolution of Paul Lansky's Music Compositions." Ph.D., University of Rochester, Eastman School of Music, 1990.

Partch, Harry, and Philip Blackburn. *Harry Partch*. Vol. 3. Enclosure. Saint Paul, MN: American Composers Forum, 1997.

Pastre, Guillaume. *Un art de la cohérence: Different Trains, Steve Reich*. Univers musical. Paris: L'Harmattan, 2018.

Patel, Aniruddh D. *Music, Language, and the Brain*. Oxford: Oxford University Press, 2008.

Perry, Jeffrey. "The Inner Voices of Simple Things: A Conversation with Paul Lansky." *Perspectives of New Music* 34, no. 2 (1996): 40–60.

Pierson, Marcelle Coulter. "The Voice under Erasure: Singing, Melody and Expression in Late Modernist Music." University of Chicago, 2015.

Prieto, Eric. "Speech Melody and the Evolution of the Minimalist Aesthetic in Steve Reich's The Cave." *Circuit: Musiques Contemporaines* 12, no. 2 (2002): 21–44.

Provenzano, Catherine. "Auto-Tune, Labor, and the Pop-Music Voice." In *The Relentless Pursuit of Tone: Timbre in Popular Music*, edited by Robert Wallace Fink, Melinda Latour, and Zachary Wallmark, 159–181. New York, NY: Oxford University Press, 2018.

Raber, Rebecca. "Charles Spearin: The Happiness Project." Pitchfork, June 1, 2009. https://pitchfork.com/reviews/albums/13089-the-happiness-project/.

Rare Billie Holiday Interview Pt. 2. Accessed September 28, 2019. https://www.youtube.com/watch?v=bENqqcFDIME.

Reich, Steve. *Writings on Music, 1965–2000*. Oxford University Press, 2002.

Remez, Robert E., Philip E. Rubin, David B. Pisoni, and Thomas D. Carrell. "Speech Perception without Traditional Speech Cues." *Science* 212, no. 4497 (1981): 947–50.

Rings, Steven. "Speech and/in Song." In *The Voice as Something More: Essays toward Materiality*, edited by Martha Feldman and Judith T. Zeitlin, 37–53. Chicago: University of Chicago Press, 2019.

Roberts, Brian, Robert J. Summers, and Peter J. Bailey. "The Perceptual Organization of Sine-Wave Speech under Competitive Conditions." *The Journal of the Acoustical Society of America* 128, no. 2 (August 1, 2010): 804–17.

Rose, Francois. "Introduction to the Pitch Organization of French Spectral Music." *Perspectives of New Music* 34, no. 2 (Summer 1996): 6–39.

Rousseau, Jean-Jacques. "Essay on the Origin of Languages (c. 1763)." In *Essay on the Origin of Languages and Writings Related to Music*, translated by John T. Scott, 7:289–332. The Collected Writings of Rousseau. Hanover: University Press of New England, 1998.

Ruwet, Nicolas. *Langage, musique, poésie*. Collection Poétique. Paris: Éditions du Seuil, 1972.

Sancheneering. "The Rick and Morty Band | Ants in My Eyes Johnson Cover on Drums, Guitar, Bass, and Sax." YouTube, September 7, 2017. https://www.youtube.com/watch?v=Pvidk3w4MfA.

Saturday Night Live. *Whatcha' Cookin' On*. New York: NBC Universal, 2020. https://www.youtube.com/watch?v=6zgSiNMSu8E&feature=youtu.be.

Schaeffer, Pierre. *Treatise on Musical Objects: An Essay across Disciplines*. Translated by Christine North and John Dack. Vol. 20. California Studies in 20th-Century Music. Oakland, California: University of California Press, 2017.

Scherzinger, Martin. "Curious Intersections, Uncommon Magic: Steve Reich's It's Gonna Rain." *Current Musicology* 79/80 (2005). https://currentmusicology.columbia.edu/article/curious-intersections-uncommon-magic-steve-reichs-its-gonna-rain/?article=curious-intersections-uncommon-magic-steve-reichs-its-gonna-rain&post_type=article

Schleifer, Ronald. *Modernism and Popular Music: Language and Music in Gershwin, Porter, Waller, and Holiday*. Cambridge: Cambridge University Press, 2011.

schmoyoho. "BED INTRUDER SONG!!!" YouTube, July 31, 2010. https://www.youtube.com/watch?v=hMtZfW2z9dw.

———. "Unbreakable Kimmy Schmidt - Songify This! (Theme Written by Jeff Richmond)." YouTube, March 6, 2015. https://www.youtube.com/watch?v=WYNbp0u8WjA.

Schoenberg, Arnold. *Fundamentals of Musical Composition*. Edited by Gerald Strang and Leonard Stein. London: Faber, 1999.

———. "[No Title]." Text manuscript. Vienna, Austria, n.d. T53.06. Arnold Schönberg Center.

———. *Verklärte Nacht and Pierrot Lunaire*. New York: Dover Publications, 1994.

Schwartz, Elliott, and Daniel Godfrey. *Music since 1945: Issues, Materials, and Literature*. New York: Schirmer Books, 1993.

Schwarz, David. "Listening Subjects: Semiotics, Psychoanalysis, and the Music of John Adams and Steve Reich." *Perspectives of New Music* 31, no. 2 (1993): 24–56.

Shintel, Hadas, Howard C. Nusbaum, and Arika Okrent. "Analog Acoustic Expression in Speech Communication." *Journal of Memory and Language* 55, no. 2 (August 1, 2006): 167–77.

Smith, Hazel. *The Contemporary Literature-Music Relationship*. Routledge Interdisciplinary Perspectives on Literature 65. New York: Routledge, 2016.

Spearin, Charles. *The Happiness Project*. CD. A&C039. Toronto: Arts & Crafts, 2009.

Sterne, Jonathan. "On the Future of Music." In *Cybersounds: Essays on Virtual Music Culture*, edited by Michael D. Ayers, 255–263. New York: Peter Lang, 2006.

———. *The Audible Past: Cultural Origins of Sound Reproduction*. Durham: Duke University Press, 2003.

Stewart, Jesse. "DJ Spooky and the Politics of Afro-Postmodernism." *Black Music Research Journal* 30, no. 2 (2010): 337–362.

Stoever, Jennifer Lynn. *The Sonic Color Line: Race and the Cultural Politics of Listening*. New York: New York University Press, 2016.

Stokes, Martin. "'Abd al-Halim's Microphone." In *Music and the Play of Power in the Middle East, North Africa and Central Asia*, edited by Laudan Nooshin, 49–61. SOAS Musicology Series. Farnham: Routledge, 2009.

———.*The Republic of Love: Cultural Intimacy in Turkish Popular Music*. Chicago Studies in Ethnomusicology. Chicago: University of Chicago Press, 2010.

Storino, Mariateresa, Rossana Dalmonte, and Mario Baroni. "An Investigation on the Perception of Musical Style." *Music Perception: An Interdisciplinary Journal* 24, no. 5 (2007): 417–32.

Tatham, Mark, and Katherine Morton. *Expression in Speech*. New York: Oxford University Press, 2004.

Tenney, James. "Meta ⊣ Hodos: A Phenomenology of Twentieth-Century Musical Materials and an Approach to the Study of Form (1961)." In *From Scratch: Writings in Music Theory*, edited by Larry Polansky. Urbana: University of Illinois Press, 2015.

———. *Song "n" Dance for Harry Partch: For Adapted Viola, Diamond Marimba, Strings and Percussion*. Lebanon, NH: Frog Peak Music, 1999.

Thieberger, Ed M., and Charles Dodge. "An Interview with Charles Dodge." *Computer Music Journal* 19, no. 1 (1995): 11–24.

Thompson, Marie. "Whiteness and the Ontological Turn in Sound Studies." *Parallax* 23, no. 3 [84] (2017): 266–82.

Tierney, Adam, Fred Dick, Diana Deutsch, and Marty Sereno. "Speech versus Song: Multiple Pitch-Sensitive Areas Revealed by a Naturally Occurring Musical Illusion." *Cerebral Cortex* 23, no. 2 (February 2013): 249–254.

Tyrrell, John. "Janáček and the Speech-Melody Myth." *The Musical Times* 111, no. 1530 (1970): 793–96.

Urban, Greg. "Ritual Wailing in Amerindian Brazil." *American Anthropologist* 90, no. 2 (1988): 385–400.

Verdyckt, Ingrid, and Dominique Morsomme. "Perception of Emotion in Impaired Facial and Vocal Expression." In *Emotions in the Human Voice*, edited by Hans Von Leden and Luigi Anolli. San Diego: Plural Publishing, 2008.

Villarreal, Yvonne. "'Kimmy Schmidt' Composer Jeff Richmond on the Show's Viral-Friendly Songs." latimes.com, March 31, 2015. https://www.latimes.com/entertainment/music/posts/la-et-ms-kimmy-schmidt-jeff-richmond-peeno-noir-20150327-story.html.

Wannamaker, Robert A. "The Spectral Music of James Tenney." *Contemporary Music Review* 27, no. 1 (February 1, 2008): 91–130.

Whitesell, Lloyd. "White Noise: Race and Erasure in the Cultural Avant-Garde." *American Music* 19, no. 2 (2001): 168–189.

Wingfield, Paul. "Janáček's Speech-Melody Theory in Concept and Practice." *Cambridge Opera Journal* 4, no. 3 (1992): 281–301.

Yeston, Maury. *The Stratification of Musical Rhythm*. New Haven: Yale University Press, 1976.

Young, Miriama. *Singing the Body Electric: The Human Voice and Sound Technology*. Farnham, Surrey: Ashgate, 2015.

Yuan, Jiahong, and Mark Liberman. "F_0 Declination in English and Mandarin Broadcast News Speech." *Speech Communication* 65 (November 2014): 67–74.

Z, Pamela. *A Delay Is Better than a Disaster*. CD. Vol. ST-214. Starkland, 2004.

———. "A Tool Is a Tool." In *Women, Art and Technology*, edited by Judy Malloy, 348–361. Leonardo. Cambridge, Mass.: MIT Press, 2003.

Zbikowski, Lawrence M. "Music, Analogy, and Metaphor." In *The Routledge Companion to Music Cognition*, edited by Richard Ashley and Renee Timmers, 501–512. New York: Routledge, 2017.

———. "Musical Coherence, Motive, and Categorization." *Music Perception: An Interdisciplinary Journal* 17, no. 1 (1999): 5–42.

———. "Musical Gesture and Musical Grammar: A Cognitive Approach." In *New Perspectives on Music and Gesture*, edited by Anthony Gritten and Elaine King, 76–87. SEMPRE Studies in the Psychology of Music. Farnham, Surrey: Ashgate, 2011.

———. *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*. AMS Studies in Music. New York: Oxford University Press, 2002.

———. *Foundations of Musical Grammar*. Oxford Studies in Music Theory. New York, NY: Oxford University Press, 2017.