Analyzing Vocal Placement in Recorded Virtual Space

Michèle Duguay

NOTE: The examples for the (text-only) PDF version of this item are available online at:
https://www.mtosmt.org/issues/mto.22.28.4/mto.22.28.4.duguay.php

ABSTRACT: When listening to a piece of recorded music through headphones or stereo speakers, one hears various sound sources as though they were in a *virtual space*. This article establishes a methodology for analyzing *vocal placement*—the apparent location of a voice within this virtual space—in recent popular music. In contrast to existing analytical approaches to the analysis of virtual space that rely primarily on close listening, I provide an empirical method using digital sound-processing tools to precisely locate recorded sound sources in virtual space. Additionally, I offer a measurable way to visually display these parameters within the virtual space. After Ruth Dockwray and Allan Moore (2010), I depict the virtual space as an empty room in which sound sources are located. I position the voice in this empty room according to five parameters: (1) *width* (the perceived lateral space occupied by the voice); (2) *pitch range* (the range occupied by the voice); (3) *prominence* (the amplitude of the voice in relationship to other sound sources in the mix); (4) *environment* (the level of echo and reverberation applied to the voice); and (5) *layering* (supplementary vocal tracks added to the voice). To illustrate the methodology, I analyze and compare the vocal placements in Eminem's and Rihanna's four collaborations: "Love the Way You Lie" (2010), "Love the Way You Lie (Part II)" (2010), "Numb" (2012), and "The Monster" (2013).

[1.1] Put on a pair of headphones and listen to Eminem's "The Monster" (**Audio Example 1**).[1] Rihanna, the featured artist, opens the song with a chorus. Her voice immediately assumes a wide physical presence as it fills the space created around you by your headphones. Only supported by an understated synthesizer, Rihanna sounds as if she's singing in a large room across which every word is reverberated. Listen to the way echoes of "head" and "bed" are dispersed within this space: a repetition of the word emerges from your left headphone, before bouncing toward the right. As the synthesizer crescendos, the first chorus ends with Rihanna's utterance of the word "crazy." Even before the word's echo fully tapers off, Eminem enters with his first verse (**Audio Example 2**). The reverberant space evoked by Rihanna's voice disappears. Accompanied by both

the synthesizer and percussion, Eminem's voice is centered, focused, and non-reverberated. Within the song's first thirty seconds, you have been sonically introduced to Rihanna and Eminem. The singer's voice reverberates across the sonic landscape, while the rapper's voice remains consistently focused.

[1.2] Upon listening to "The Monster," you are perceiving a virtual space—an imaginary physical location molded by reverberation, stereo placement, and various other sonic parameters. Auditory cues in the recording contribute to your impression of a place in which you, the listener, are located. According to Eric Clarke, a virtual space is a setting "specified by the same perceptual attributes as a real space, but which is not physically present at the time" (2013, 95). A listener, in other words, relies on analogous aural cues to understand virtual spaces and real spaces. A recorded performance can metaphorically position this listener in a virtual space that evokes a large concert hall, a crowded room, or an otherworldly space. Virtual space, then, is the abstract space in which a recorded performance unfolds.

[1.3] Virtual spaces are not necessarily meant to emulate the physical location in which a performance was recorded. In popular music, for instance, virtual spaces emerge as the result of a series of technical and creative decisions by producers, engineers, and artists. Before reaching your ears, Rihanna's and Eminem's original performances have been mediated by several effects—added reverberation, panning, EQ, compression, and so on. The virtual space heard in "The Monster," then, is not an exact sonic reproduction of the location in which Eminem and Rihanna were recorded. Manipulated versions of Rihanna and Eminem's voices as they unfold in an artificial space are presented to our ears after a series of creative decisions made in the studio. Indeed, Simon Zagorski-Thomas writes that a sound engineer can "guide us toward hearing in a particular way," asserting that record production can be studied as a set of techniques "that seek to manipulate the way we listen" (2014, 72). As Ragnhild Brøvig-Hanssen and Anne Danielsen point out, such virtual spaces can aim to emulate real-life environments while others "[cultivate] spatialities that do not exist outside of technologically mediated environments"—that is, they are imaginary locations that could not exist in the physical world (2016, 22). Virtual spaces are carefully crafted components of recorded popular music, and as integral a part of the creative product as pitch, rhythm, and form. These spaces, which shape how one perceives and relates to recorded sound sources, are one of the central parameters through which a listener experiences and relates to recorded music.

[1.4] In this article, I am interested in analyzing the way voices unfold in virtual space. In *A Blaze of Light in Every Word*, Victoria Malawey writes that "for many genres, vocal delivery remains the single most common compelling feature . . . In short, voice matters in pop" (2020, 2–3). Indeed, several recent analytical interventions in the study of popular music analysis focus on the singing voice. Kate Heidemann, for instance, outlines a method for analyzing the embodied aspects of vocal timbre (2014, 2016). Malawey's extensive study proposes analytical tools for precisely discussing various facets—timbre, quality, prosody—of the recorded voice in pop. As a complement to studies of the vocal timbre and performance, this article provides a method for analysts to describe how voices are positioned within virtual space. Such a method allows analysts to explain *how* a vocal performance is presented to listeners. Consider the following questions: (1) How can we describe the way Rihanna's and Eminem's voices unfold in virtual space throughout "The Monster"? (2) How can we compare the way their voices are spatialized throughout the song? This article provides tools for answering such questions. I propose a methodology for analyzing *vocal placement*—the location of a voice within a virtual space.[2] The methodology allows analysts to precisely locate, visualize, and compare the ways in which different voices are positioned in a recording. I therefore equip analysts with a reproducible methodology that allows for the study of a salient parameter of popular music.

[1.5] The article begins with a review of analytical approaches to virtual space. Several authors have approached the study of virtual space—with or without a focus on the voice—from a variety of angles. While existing methodologies provide useful starting points for analyzing recorded space, I argue for the necessity of two interventions: (1) a consensus on the parameters that contribute to virtual space, and (2) the development of a precise method for describing and

comparing the way in which voices are spatialized in recordings. I then outline a methodology for analyzing vocal placement in a recorded track. Not only does the proposed methodology provide a clear set of parameters for discussing vocal placement in virtual space, but it also allows for direct comparisons between the way voices spatially unfold across different recordings. I depict virtual spaces as three-dimensional plots containing individual sound sources (after Dockwray and Moore 2010). I posit that five parameters contribute to vocal placement within this virtual space: width, pitch range, prominence, environment, and layering. I then provide a description of each parameter, along with instructions for analyzing them. The methodology outlined in this article allows analysts to precisely engage with a sonic parameter of recorded music that is central to the listening experience. Finally, I showcase the methodology by analyzing and comparing the vocal placements in Eminem and Rihanna's four collaborations: "Love the Way You Lie" (2010), "Love the Way You Lie (Part II)" (2010), "Numb" (2012), and "The Monster" (2013).

[1.6] This article outlines the analytical methodology with a central focus on technical and financial accessibility. When I began familiarizing myself with literature some aspects of computer-assisted music analysis for this project, I experienced a steep learning curve due to the significant number of specialized terms. To make the methodology as user friendly as possible, I aim to provide clear definitions for terminology related to computer-assisted audio analysis. In addition, my methodology relies exclusively on free and open-access tools: source-separation algorithm Open-Unmix, Sonic Visualiser, and panning visualization tool MarPanning. Analysts do not require any access to costly software, or to additional resources that may be rare, expensive, or difficult to obtain (a recording studio, original recording session files, or DAW project files). By prioritizing technical and financial accessibility, I hope to make the analytical method readily available to a wide range of music researchers.

[1.7] In "Specters of Sex" (2020), Danielle Sofer encourages music theorists to consider and rethink the audience intended for their work. Music theories are technologies, they argue, and technologies have users. These users are often assumed to be white, cisgender, and heterosexual men. Sofer writes: "If instead we explicitly name women—who may also be black scholars, scholars from Asia or the Global South, and/or members of the LGBTQIA+ community—as part of the users of our theories, our music-theoretical objectives may very well change" (2020, 56). Following Sofer's exhortation, I want to explicitly invite scholars of marginalized genders to make use of the analytical methodology outlined here. While women such as Anne Danielsen, Ruth Dockwray, Nicola Dibben, and Kate Heidemann have made significant contributions to this area of research, a reader may nonetheless notice that most analysts who work on virtual space are men. Moreover, this area of inquiry is overwhelmingly white. As the study of virtual space and vocal placement develops, I hope to see further non-white and non-male-led developments in this area of research. Finally, I also hope that analysts will use this methodology as an invitation to engage in interdisciplinary work that helps us rethink the objectives of music theory. As I discuss elsewhere, for instance, vocal placement can be a potential tool for unpacking the different ways in which gender is created through sound (Duguay 2021). To hint at such future work, my analysis of the four Rihanna/Eminem collaborations ends with a brief reflection on how vocal placement contributes to the sonic formation of gendered difference in recorded popular music collaborations.

Analytical Approaches to Virtual Space

[2.1] In the past three decades, several studies have offered analytical approaches to virtual space. [3] Rather than focusing solely on the voice, these studies tend to address virtual space more generally through a consideration of all recorded sound sources and instruments present in a recording.[4] Peter Doyle's (2005) study of American popular music from 1900 to 1960 is one of the first analytical studies of virtual space. Through an in-depth study of monoaural recordings, Doyle asserts that the application of echo and reverberation can reinforce the lyrical aspects of a song. He considers for instance the groundbreaking use of slap-back echo in Elvis Presley's 1956 major-label debut "Heartbreak Hotel." The echo sonically positions Presley in a wide, empty space, compounding the feeling of loneliness conveyed by the lyrics.

[2.2] Aside from Doyle's foray into monoaural recordings, most studies on virtual space analyze stereophonic sound. Ruth Dockwray and Allan Moore's sound-box (2010) is a key example of such a method.[5] In essence, the sound-box looks like an empty room. It represents the virtual space experienced by a listener upon hearing a recorded piece of music. An analyst can display the location of various sound sources—instruments, voices, and sound-effects—within this space. Consider for instance the sound-box illustration of Jimi Hendrix's "Purple Haze" (**Example 1**). The image shows three parameters:

1. The lateral placement of a sound within the stereo field (left-right axis). Hendrix's voice—depicted as floating vocal folds—originates from the right side of the space. Drums, guitar, and bass are located at the center of the mix.

2. The proximity of a sound (front-back axis). As the loudest sound of the track, Hendrix's voice hovers toward the front. The remaining instruments, which are quieter, are relegated to the back.

3. The height of a sound according to frequency (high-low axis). Notice how the bass guitar, with its low range, floats right above the floor. Hendrix's high-pitched voice, conversely, is closer to the ceiling.

The sound-box can also express a fourth, and final, parameter, which is absent from "Purple Haze": The movement of sounds through time. A guitar moving from the left to the right channel, for instance, could be depicted by an arrow or an animation.

[2.3] In his book *Rock: The Primary Text*, Moore first outlined the idea of the sound-box, which he "envisaged as an empty cube of finite dimensions, changing with respect to real time" (1993, 106). Around the same time, Danielsen independently developed an analogous "sound-box" (1993, 1997).[6] In order to analyze Prince's recordings, she wrote, "it seems useful to think of sound as a three-dimensional sound-box, with front-rear, left-right and high-low along the three axes. The positioning along the front-rear axis depends on a combination of two parameters: timbre and dynamics. Left to right refers to the mix, while high to low is tied up to pitch and frequencies" (1997, 283). Danielsen and Brøvig-Hanssen further developed this concept in their book Digital Signatures (2016). The parameters used by Danielsen to delineate her version of the sound-box are analogous to those of Dockwray and Moore. The primary difference between their methodologies, however, is that Dockwray and Moore rely primarily on mapping sounds through visual representations while Danielsen (subsequently with Brøvig-Hanssen) models the listening experience through verbal descriptions.[7] Authors that cite Danielsen's methodology tend to follow suit. In her study of Nina Simone samples in hip-hop music, for instance, Salamishah Tillet uses Danielsen's theorization of the sound-box to provide in-depth written descriptions of the placement of Simone's voice throughout Cassidy's "Celebrate" (2014).

[2.4] In his doctoral dissertation, Serge Lacasse (2000) proposes another approach to the visual representation of recorded space. Focusing exclusively on the voice, he coins the term "vocal staging" in reference to "any deliberate practice whose aim is to enhance a vocal sound, alter its timbre, or present it in a given spatial and/or temporal configuration" (2000, 4).[8] These practices include manipulations in (1) environment (echo, reverberation); (2) stereo image (lateral location in space); and (3) distance (changes in intensity). Phasing, saturation, flanging, and distortion also contribute to vocal staging. Lacasse displays vocal staging through maps showing any enhancements, manipulations, and effects applied to the voice as it evolves through time (**Example 2**). The first row of the example displays approximately the first minute of the track. Lacasse provides timestamps at the onset of each significant vocal staging event, each of which is labeled with a number (S1, S2, etc.) At 00:05, for instance, a reverberation effect (S2) begins. In a 2010 article, Lelio Camilleri proposes an alternative way to visually represent virtual space. He considers three features of a recording: (1) localized space (the depth and position of a sound within the stereo field); (2) spectral space (the timbral saturation of a sound, or, the extent to which a sound fills up the spectral space) and (3) morphological space (the interaction between localized space and spectral space). Camilleri illustrates the first feature, localized space, through two-dimensional diagrams called "Stereo Windows" (**Example 3**). The vertical dotted line represents the center of the stereo stage, while the spaces on either side of the line represent the left and the right channels. Camilleri shows the location of instruments in the mix by arranging them on this two-dimensional plane: the bass, for instance, begins in the right channel and eventually moves

through the left. This motion is indicated by an arrow. The methods designed by Lacasse and Camilleri allow for a great degree of specificity in showing the individual elements of virtual space as they evolve through time. Their proposed visual representations, however, flatten the virtual space to a two-dimensional plane that does not capture the listening experience as intuitively as the three-dimensional sound-box representations.

[2.5] Some analysts are especially interested in how virtual space interacts with other musical parameters. Zachary Zinser (2020), for instance, explores how virtual spaces intersect with texture and form in recent popular music.[9] For Victoria Malawey (2020), virtual space is one of the central components of vocal delivery in recorded popular songs. She outlines a framework for analyzing a voice's "Mediation with Technology," including aspects of reverb, delay, and stereo panning. Other authors study the impact of virtual space on a song's theme, atmosphere, or lyrical content. Brøvig-Hanssen and Danielsen (2016), for instance, consider the ways in which of echo and reverberation create the impression of an otherworldly space in Kate Bush's "Get Out of My House" (1982). Moore, Schmidt and Dockwray (2011) identify four zones in which a track can unfold: intimate, personal, social, and public.[10] This framework allows them to explore how different spaces convey musical meaning. In "Standing in the Way of Control" (2016), Beth Ditto's "overdriven voice"—loud, aggressive, and prominent—rejects the regulated femininity afforded to white Western women.[11] Finally, in her extensive study of vocal timbre in popular music, Kate Heidemann counts virtual space as one of the parameters through which singers can express gendered and racialized subjectivities (Heidemann 2014). She shows for instance that the recorded performance of Loretta Lynn's "Fist City" (1968) uses panning to emulate instrument placement on a real stage. This evokes a working-class performance venue with little to no amplification and mixing equipment, where Lynn sings about gender roles in a way that evokes a "rugged working-class model of femininity" (Heidemann 2014, 31).

[2.6] Other scholars (Gamble 2019; Greitzer 2013; Kraugerud 2020; and Malawey 2020) have also offered analyses tying aspects of a song's perceived virtual space with notions of meaning and identity. Such studies often feature rich close readings of individual pieces that allow the analyst to evocatively interpret the way virtual space relates to lyrics, vocal delivery, and/or a vocalist's identity. Lacasse and Mimnagh (2005) and Burns and Woods (2004), for instance, explore how Tori Amos's intimate, enclosed vocal performance in her cover of Eminem's "97 Bonnie and Clyde" radically alter the meaning of the song's lyrics. While this analytical literature is central to my understanding of the importance of virtual space, I depart from such close readings in this article presenting a somewhat more positivist approach that relies on computer-assisted analysis. My methodology for analyzing vocal placement complements these studies by proposing a precise way to quantify and measure the way voices operate within recorded spaces. As expressed earlier in the article, however, I hope that the methodology will be used to ask analytical questions about sound and identity.

[2.7] Through a variety of approaches, the authors discussed above advocate for virtual space to be taken seriously as a musical parameter. This body of literature provides valuable ways to analyze virtual space as it interacts with other musical parameters. From these studies, however, emerge two issues that warrant further discussion. First, there is a lack of consensus among the authors about the sonic parameters that constitute virtual space. Second, while these methodologies allow an analyst to describe various virtual spaces, they do not explicitly provide a reliable way to *compare* the virtual spaces of different recordings.

Issue 1: What sonic parameters contribute to the formation of virtual space?

[2.8] At first glance, the works reviewed above seem to discuss the same phenomenon—virtual space. Each author, however, constructs their methodology around different elements of a recording. As a result, there is no consensus on the sonic parameters that constitute virtual space. These discrepancies are illustrated in **Example 4**.[12] Notice, for instance, how Danielsen, Camilleri, and Malawey are the sole analysts who explicitly name timbral saturation as a central component of virtual space. Conversely, all analysts—except Doyle, who is studying monaural recordings—consider lateral placement as a parameter of virtual space. A subtler issue arises when two or more

authors discuss the same parameter, but with divergent definitions and conceptualizations. The notion of "distance"—occasionally called "depth"—is a typical case. For Camilleri, distance the sense of depth is created through reverberation. The more reverberant the virtual space, the more depth is perceived by the listener. Camilleri is not referring to the distance between sound source and listener, but to the size of the space in which these sound sources unfold.[13] For Lacasse, distance refers to the position of a sound along the depth axis of the virtual space. Unlike Camilleri, he does not consider reverberation the primary factor in creating distance. Rather, he argues that the illusion of distance is created by a combination of timbral parameters, reverberation, loudness, and microphone technique. Finally, for Dockwray and Moore, a listener perceives the distance of a sound via a combination of volume and reverberation. Three different studies on the creation of virtual space, then, offer three different definitions of the same concept. These inconsistencies raise several questions: What, exactly, is distance? Does it refer to the size and dimensions of the virtual space, the perceived gap between listener and sound source, or the perceived placement of the listener within the virtual space? What sonic parameters participate in its creation?

Issue 2: How can we precisely compare virtual spaces?

[2.9] Suppose that you want to compare the virtual space in "The Monster" with the virtual space created in "Love the Way You Lie" (2010), another collaboration between Eminem and Rihanna. Which space is more reverberant? Is Eminem's voice more salient in the first or second song? How similar are the voice's locations on the stereo stage? In order to answer these questions, you would need to precisely quantify each space's characteristics (i.e., "the voice is panned at 30% to the right"). Current methodologies, however, do not provide a precise way to describe virtual spaces. As a result, different virtual spaces cannot accurately be compared. Consider the following two analytical statements. First, Dockwray and Moore write that in "Congratulations" (1968) by Cliff Richard, "the brass is more prominent and brought forward in the mix" (2010, 189). Second, in his analysis of Alanis Morissette's "Front Row" (1998), Lacasse writes that "Interestingly, the 'God' setting consists of a very high level of reverberation with a quite long reverberation time, typical of very large environments such as concert halls, caves, or cathedrals" (2000, 227).

[2.10] Both claims are based on close listening of the tracks: Dockwray and Moore hear the brass as more prominent than other instruments, and Lacasse hears reverberation and associates it with spacious environments. Such personal readings are especially useful when making claims about a single track: Dockwray and Moore are interested in the diagonal mix in "Congratulations," and Lacasse ties the reverberation in "Front Row" to word painting. These claims, however, are seemingly unsupported by any explicit methodology, procedure, or instructions on how to listen for and describe virtual spaces. Without such guidance, a reader cannot necessarily know for certain that they are attuned to the same musical parameters that the author is describing. A lack of a clear methodology or procedure also becomes problematic when attempting to compare two or more tracks. Consider the following questions: Is the brass more prominent in "Congratulations" than in Tower of Power's "What Is Hip" (1973)? Does "Front Row" evoke the widest space in Morissette's album *Supposed Former Infatuation Junkie*? Both questions require us to compare two or more virtual spaces. This cannot be achieved, however, without a clear way to measure parameters like "prominence" and "reverberation." An analyst might be able to hear varying degrees of reverberation, for instance, but would need a precise scale on which to locate the different degrees. [14] While comparisons of vocal placements are helpful from a purely technical perspective, they are also crucial for an analyst wishing to make claims about the way vocal placement interacts with other aspects of the music, such as vocal timbre or the way a vocal performance expresses identity through sound.

[2.11] As a response to the two limitations described above, my method for analyzing vocal placement within virtual space features: (1) a clear definition of the sonic parameters contributing to the perception of space; and (2) a reproducible, step-by-step methodology allowing for consistent and replicable descriptions of various virtual spaces. In addition to supporting close readings of individual pieces, this analytical tool enables reliable comparisons of virtual spaces by presenting a series of five-point scales on which specific sonic parameters can be located.

Analyzing Vocal Placement in Virtual Space

[3.1] Borrowing Dockwray and Moore's visual representation of the sound-box, I depict the virtual space as an empty three-dimensional room in which sound sources are located. The analyses outlined throughout this article assume the existence of a listener who is positioned in front of the virtual space, at the location indicated by an X (**Example 5**). A listener perceives this virtual space as though it were a stage on which a virtual performance was unfolding.[15] In the analyses presented throughout the article, the listener in question is myself—the analyses capture my own listening experience as I listen to recorded performances via headphones.[16] Five parameters contribute to vocal placement within the virtual space: (1) width; (2) pitch range; (3) prominence; (4) environment; and (5) layering. The first three parameters correspond to the three axes of the three-dimensional sound-box. Like Dockwray and Moore (2010), Brøvig-Hanssen and Danielsen (2016), and many others, I equate width with the horizontal axis and pitch with the vertical axis. In a departure from other methodologies, I equate the depth axis with prominence. The environment parameter encompasses both echo and reverberation, which are two temporal effects. Finally, while layering is not typically included in discussions of virtual space, I discuss it to show how additional vocal layers—which are a crucial component in many recorded vocal performances—are placed within the virtual space. The inclusion of layering in the methodology allows me to account for the spatialization of all aspects of a recorded voice.

[3.2] Some of the parameters discussed in the article are perceived through the same auditory cues that one uses in the "real world." Environment, layering, and width, for instance, rely on auditory phenomena—echo, reverberation, and lateral placement—that (are used in everyday life). The methodology requires analysts to use these real-life perceptual cues to evaluate the way a voice is positioned in vocal placement. As explained below, however, the remaining two parameters rely on a more metaphorical understanding of space. Pitch range is mapped on the top-down axis according to frequency, for instance, while prominence is a measurement proposed as an alternative to distance, a parameter that can only be accurately perceived in "real" life. The virtual space as it is modeled in this article therefore relies both on real-life perceptual cues and on the metaphorical spatialization of musical parameters. This tension is a reminder that virtual spaces are imagined environments brought to life in a listener's mind, not exact replicas of real-life locales.

[3.3] For each parameter, I provide a definition and an explanation of the parameter's significance to vocal placement within virtual space. The vocabulary introduced in these discussions is intended to provide music analysts with specific adjectives to describe parameters of vocal placement. Finally, I provide instructions for visually displaying the parameter in the virtual space. The resulting visual representations allow an analyst to succinctly depict the way a voice unfolds in the abstract space created in the recording.

[3.4] As shown in **Example 6**, the methodology relies on both computer-assisted audio feature extraction and aural analysis. An isolated vocal track is first extracted from the full mix. This isolated track allows the analyst to extract information specific to the placement of the voice, without having to consider additional sound sources—guitars, bass, percussion, and so on—in the virtual space. Data on pitch range and prominence is extracted from audio files through sound analysis software Sonic Visualiser. Pitch (expressed in Hz or with pitch names) and prominence (expressed as a percentage) values are continuous data, as opposed to discrete.

[3.5] Width is analyzed by ear with the aid of stereo panning visualization tool MarPanning. Finally, the environment and layering parameters are determined completely by ear. The width, environment, and layering parameters are accompanied by five-point Likert-type scales meant to guide the analyst's listening.[17] The width, environment, and layering parameters are expressed as categorical data, because there is a finite number of values for each parameter. Consider, for instance, the scale for analyzing layering. As will be discussed later, a voice can be sorted into one of five layering categories: L1, L2, L3, L4, or L5. Each category represents a different scenario under which a voice can operate in a recorded environment. Additionally, the data for width, environment, and layering are ordinal because they are ranked in a specific order: L1 and L5, for instance, are the values with "the least" and "the most" layering.

[3.6] To conduct an analysis of a voice's placement in virtual space, an analyst requires the following materials:

1. a .wav file of the song under study;

2. an isolated vocal track (an altered version of the song containing the vocals only);

3. access to Sonic Visualiser;[18] and

4. access to panning visualization tool MarPanning (McNally, Tzanetakis, and Ness 2009).

Creating an Isolated Vocal Track

[3.7] To isolate the vocal track from a recorded song, I recommend using source separation algorithm *Open-Unmix*. Open-Unmix is a "state of the art open-source separation system" (Stöter et al. 2019, 4). Open-Unmix allows users to separate a complete mix into four separate tracks: (1) Vocals; (2) Drums; (3) Bass; and (4) Other.[19] The algorithm has no graphic user interface (GUI), a potential barrier to users with no experience with coding language Python. Open-Unmix is nonetheless rife with advantages: it's free and allows for the simultaneous separation of multiple audio files. Some commercially available sound editing software, like Audionamix XTRAX STEMS and iZotope's Music Rebalance feature, also offer tools for isolating vocals from a full mix. These tools have the advantage of presenting user-friendly GUIs, but they are significantly expensive.

[3.8] Audio Examples 3 to 5 shows three versions of the first chorus of "Love the Way You Lie" by Eminem feat. Rihanna. The first version is the full, unaltered track (**Audio Example 3**). The second (**Audio Example 4**) and third (**Audio Example 5**) versions are isolated vocal tracks created with iZotope RX 7 and Open-Unmix, respectively. Note that the isolated vocal tracks are not perfect: they often contain audio artifacts from other sound sources. The isolated track created by iZotope, for instance, still contains the piano notes that overlap in pitch with Rihanna's sung melody. In the Open-Unmix version of the isolated vocal track, the piano is almost completely absent, but some aspects of the voice, such as the low amplitude reverberated images, are not captured by the source separation algorithm and therefore not included in the isolated vocal track. If one can choose between multiple isolated tracks when analyzing vocal placement, I would recommend using the track that captures all aspects of the recorded voice as accurately as possible. In this case, for instance, I will use the iZotope version for my analysis (Audio Example 4).

Analyzing the Width of a Voice

[3.9] The lateral axis of the virtual space represents the stereo stage.[20] The **width** of a voice is the breadth it occupies on this stereo stage. Sounds occupying a narrow space can be described as localized (i.e., "the voice is localized in the center of the virtual space"), while wider sounds can be described as diffuse (i.e., "the guitar sound is diffuse as it is spread throughout the stereo stage"). The width of a voice is displayed in the virtual space through the breadth of the image: localized sounds look narrow and diffuse sounds look wide (**Example 7**).

[3.10] An analyst can determine the width value of a voice through aural analysis supported by sound processing tool MarPanning (McNally, Tzanetakis, and Ness 2009). MarPanning provides a visualization of how recorded sounds are panned on the stereo stage (**Example 8**).
Video Example 1, for instance, shows the MarPanning visualization of the isolated vocals in the first chorus of "Love the Way You Lie." For the most part, Rihanna's voice occupies a narrow and centered position in the virtual space. Aspects of her voice—the echoes on the word "burn," for instance—are spread across the stereo stage. Upon listening for a voice's width with the aid of MarPanning, an analyst can use the five-point scale provided in **Example 9** to determine a voice's width value. The lowest value, W1, is extremely localized; the highest value, W5, is the most diffuse. Rihanna's voice in Video Example 1, for instance, corresponds with the category W4. Example 9 also indicates a way to visually display each width value in the virtual space.

Analyzing the Pitch Range of a Voice

[3.11] The vertical axis of the virtual space represents pitch. The "floor" of the virtual space corresponds to C0 (16.35 Hz), while its "ceiling" corresponds to C7 (2093 Hz) (**Example 10**).[21] The **pitch range** of a voice is the frequency span it occupies. The pitch range of a voice is displayed by the length its image occupies on the vertical axis. Different pitch ranges can be described according to the high/low spectrum (i.e., "song X has a higher pitch range than song Y").

[3.12] An analyst can determine this pitch range through aural analysis by identifying the highest and lowest pitches performed by a vocalist in a given excerpt. In "Love the Way You Lie," for instance, Rihanna sings a melody that ranges from B♭3 to D5. For corpus studies, and in other situations in which an analyst might require pitch range data for multiple recordings, I suggest using Sonic Visualiser and its associated plug-in PYIN Smoothed Pitch Track[22] to determine the pitch range of a voice. Using automated pitch analysis allows for quicker processing of one or more recordings.

Analyzing the Prominence of a Voice

[3.13] The depth axis of the virtual space represents amplitude.[23] The prominence of a voice is its perceived amplitude in relationship with other sound sources in the recording. The prominence of a voice is expressed as a percentage, and visually displayed by its location on the depth axis. The "back wall" of the virtual space corresponds to 0% prominence while the "front wall" corresponds to 100% prominence (**Example 11**). Sounds closer to the front wall can be described as salient (i.e., "during the verse, the voice has a prominence value of 90% and is therefore salient in the virtual space"), while sounds closer to the back wall can be described as blended (i.e., "the voice in the chorus is much more blended, with a prominence value of 20%").

[3.14] An analyst can determine a voice's prominence with the following formula:

$$\text{Prominence} = \left(\frac{\text{Average RMS Amplitude of the isolated vocal track}}{\text{Average RMS Amplitude of the full track}} \right) \times 100$$

An analyst can use Sonic Visualiser and its associated plug-in RMS Amplitude to determine the prominence of a voice in the virtual space.[24] For example, in the chorus of "Love the Way You Lie," Rihanna's voice has an average RMS Amplitude of 0.06437326.[25] The full track has an average RMS amplitude of 0.07539027.[26] The prominence value of her voice is therefore as follows:

$$\text{Prominence} = \left(\frac{0.06437326}{0.07539027} \right) \times 100$$
$$\text{Prominence} \approx 85.5$$

Rihanna's voice occupies approximately 85.5% of the chorus's amplitude, creating an especially salient sound.

[3.15] My decision to equate the depth axis with amplitude warrants some discussion. The depth axis of the virtual space is somewhat a point of contention, with various analysts tying it to different parameters.[27] If virtual spaces were completely analogous to physical, real-life spaces, we could simply tie the depth axis to distance. The more distant a sound, the further it would be towards the "back wall" of the space. In virtual space, however, the sonic parameters that would normally affect our perception of distance are of no help.

[3.16] In the real world, humans can perceive aurally distance through two parameters: (1) timbral detail; and (2) the ratio of direct sound to reflected sound. In *Understanding and Crafting the Mix*, William Moylan (2015) writes that humans primarily determine distance through timbral detail. The better we perceive a sound's partials, the closer we are to the sound source. With increasing distance, our ears have more difficulty perceiving high-frequency partials from a sound since they typically have lower-amplitudes. Moylan then moves on to reverberation, writing that "as the [sound] source moves from the listener, the percentage of direct sound decreases while the percentage of reflected sound increases" (2015, 30). To picture this effect in action, imagine that you are in a (real-life) enclosed, reverberant space. Someone located near you sings a melody. You hear

the direct sound—the person's voice—along with its reverberated images—the echoes of the voice —bouncing across the room. Now, imagine that this person walks to the opposite side of the room, further away from you, and sings the same melody. The direct sound is now murkier, but the reverberation is more pronounced. These variations in the relationship between a sound and its reverberation allow you to perceive distance.

[3.17] Let us now return to virtual space. When editing and mixing a song, engineers can modify sound sources to alter our perception of distance.[28] Albin Zak (2001, 155) explains how effects of distance can be manipulated by a mixing engineer through loudness, frequency spectrum, equalization, compression, and stereo placement. Some of these parameters, however, cannot be analyzed directly from a finished mix. Consider for instance, the role of timbral detail in distance perception. Upon hearing a guitar sound in a recorded mix, we are unable to immediately pinpoint its apparent distance. To do so, we would need to hear the guitar several times at a variety of distances to gauge how distance affects its timbral level. Reverberation provides no additional help in locating the guitar sound: there is no way for us to determine how distance affects the ratio of direct to reverberant sound if we only have the finished recording. The lack of information on the way distance affects the sound we hear prevent us from reliably identifying its distance.[29] A finished recording does not provide enough information to precisely infer the virtual "distance" of a sound object in relationship to the listener. I believe that this difficulty is partially responsible for the lack of consensus about the depth axis that exists among scholars of virtual space.

[3.18] As an alternative to distance, I posit that it is the *prominence* of a sound, not its perceived distance from the listener, that determines its location on the depth axis. This framework is analogous to the sound-box, in which the front-back axis is partially defined by volume. Unlike some other analysts, however, I do not consider reverberation, delay, or echo as relevant to the depth axis of the virtual space. This makes the placement of a voice along the depth axis much easier to measure. Moreover, while these parameters are relevant to the real-life perception of distance, I believe that this is not the case in the virtual space. A highly reverberated sound does not necessarily sound "far away" or less prominent. My methodology, instead of defining distance through a mixture of sonic parameters, puts the front-back axis under the purview of amplitude only. This shift in framework allows us to discuss reverberation under independent terms that better encapsulate its role in a mix.

Analyzing the Environment of a Voice

[3.19] The environment parameter considers both reverberation and echo. Reverberation—the persistence of a sound after it is produced—occurs when a sound signal is reflected off a variety of surfaces. When these reflections are audible as distinct repetitions of a sound, we refer to echo (or delay).[30] Every conceivable indoor or outdoor space has its distinctive reverberation profile, with a unique configuration of walls, floors, surfaces, and materials that absorb or reflect sound. As such, reverberation and echo are not characteristics of the sound source itself but *characteristics of a sound's environment*.

[3.20] In commercially produced popular music, various types of artificial echo and reverberation profiles are routinely applied to recorded voices. There are at least three levels of reverberation to consider. The first level consists of the properties of the environment in which the sound was recorded. Imagine, for instance, that you record your singing voice in an empty room. The microphone will capture both the sound of your voice and its reaction to the room.[31] A second level of reverberation might be added to your recorded voice by a mixing engineer. A very short reverberation time—under 30ms, for instance—could give the recording a more present sound (Huber 2018, 73), or an added layer of echo might add a dynamic rhythmic component to the melody. Every sound source in a mix can exist in its own reverberant environment: the guitar accompanying your voice could be mixed with a contrasting reverberation profile. Finally, an additional layer of reverberation or may be applied to the entire mix.

[3.21] Virtual spaces are therefore a compound of varied sound sources placed in different environments—each with its own reverberation and/or echo profile. Moylan makes a distinction

between the "individual environments"—the qualities of the space in which a single sound source is located—and the "perceived performance environment"—the entire virtual space, which is a composite of several individual environments that may have subsequently be processed as a group (Moylan 2009). Gibson (2005) and Brøvig-Hanssen and Danielsen (2013) illustrate these nested environments by displaying smaller spaces nested within the larger-scale virtual space (**Example 12**).

[3.22] While some authors (Camilleri 2010; Dockwray and Moore 2010) link reverberation and echo with depth or distance, my model does not tie these parameters with a sound's apparent position in the virtual space. In the model of virtual space outlined in this article, location is determined through the lateral, vertical, and depth axis, which respectively correspond to panning, pitch, and prominence. These three parameters, crucially, operate independently of reverberation and echo: a reverberant sound could be placed toward the front or back of a virtual space, for instance, depending on its amplitude. In my model, reverberation and echo are instead represented through the "environment" parameter.

[3.23] The environment is the space in which a given sound object reverberates. In a virtual space, each sound object—a voice, a guitar, percussion, etc.—can have its own environment. Virtual spaces are therefore a compound of smaller-scale environments, each containing their own sound object. The environment of a voice is therefore the space in which this voice reverberates. The environment of a voice is shown through opacity: as reverberation or echo increases, opacity decreases. **Example 13** displays five different possible environment categories, along with their associated visual representation: Level E1 (completely opaque) has no reverberation, and Level E5 (very transparent) is extremely reverberant. Highly reverberant environments can be described as wet (i.e., "the voice is situated in a wet environment"), while less reverberant environments are described as dry (i.e., "the environment in the chorus is drier than that of the verse").[32]

[3.24] An analyst can measure the voice's environment through aural analysis. Through close listening, an analyst can classify a voice's environment into one of the five categories shown in Example 13. These five categories are meant to encapsulate the most common echo and reverberation tropes heard in post-2000 Top 40 music.[33] In the first chorus of "Love the Way You Lie," for instance, Rihanna's voice is treated with slight reverberation at the end of various musical phrases. According to the scale in Example 13, her voice therefore has a value of E2 for this section. Note that the environment of a voice might change over time, and the analyst should choose the level that best encapsulates the overall sound of the section under study. As explained above, each sound source in a recording might be placed in its own environment: a voice treated with a lot of reverberation, for instance, may be accompanied by a completely dry guitar. Therefore, one should ensure that their aural analysis considers only the voice and not the apparent environment of the other sound sources in the mix.

Analyzing the Layering of a Voice

[3.25] In virtual space, layering refers to the additional vocal tracks that are dubbed over a single voice.[34] More specifically, Malawey defines layering as "the generic process of playing more than one vocal track simultaneously, whether in counterpoint, harmony, or unison with the lead vocal track" (2020, 131). An unlayered vocal track can be described as single (i.e., "her voice unfolds as a single track during the verse") while multiple simultaneous tracks should be referred to as layered (i.e., "the layered vocals in the chorus contrast with the verse")[35].

[3.26] An analyst can measure the voice's layering through aural analysis. Through close listening, an analyst can classify a voice's layering into one of the five categories shown in **Example 14**. In the chorus of "Love the Way You Lie," for instance, Rihanna's voice consists of a single layer—there are no readily audible layers added to her voice, giving this chorus a layering level of L1. Note that a given layering rating might not be representative of the entirety of a given formal section. A singer might open a chorus at a layering level of L1, but quickly be joined by several added layers for a level of L4. The analyst should choose the level that best encapsulates the overall sound of the section under study.

[3.27] The previous discussion has introduced the five parameters that contribute to the placement of a voice within virtual space. The position of the voice within virtual space can be intuitively displayed through the visual conventions outlined for each parameter. Each component of a voice's vocal placement, moreover, can be described alongside a specific spectrum: the width of a voice is localized/diffuse, the pitch range is high/low, the prominence is salient/blended, the environment is dry/wet, and layering is single/layered. **Example 15** displays the Rihanna's vocal placement in the first chorus of "Love the Way You Lie." The example shows the precise location of Rihanna's voice within the virtual space, along with the aspects of reverberation and layering that also contribute to vocal placement. The W3 value of Rihanna's voice is showed via the arrows that emerge from the central image of her voice. Her voice's pitch height—G3 to D4—is indicated by the position on the vertical axis, and her voice's prominence—85.5%—is indicated by its position on the depth axis. The slightly transparent vocal image indicates the E2 environment value of her voice, while the single vocal image indicates that her voice is single (L1).

Vocal Placement in Rihanna and Eminem's Four Collaborations

[4.1] I opened this article with a brief verbal description of Rihanna's and Eminem's vocal placements in "The Monster." With the methodology outlined above, I can now more precisely describe the way both voices unfold in virtual space. The measurements obtained through my analysis also allow me to precisely compare both artists' vocal placements, both within and across tracks. Section 4 of this article, then, provides vocal placement data for Eminem's and Rihanna's four collaborations. After an overview and comparison of the way their voices are spatialized in the recording, I share some preliminary thoughts for further projects that take vocal placement as a point of departure.

[4.2] In the early 2010s, Rihanna and Eminem collaborated on four songs: (1) "Love the Way You Lie" (released in 2010 on Eminem's album *Recovery*); (2) "Love the Way You Lie (Part II)" (released in 2010 on Rihanna's album *Loud*); (3) "Numb" (released in 2012 on Rihanna's album Unapologetic); and (4) "The Monster" released in 2013 on Eminem's album *The Marshall Mathers LP 2*). I take these four collaborations as a case study to show how the methodology can be used to compare vocal placements. By analyzing width, prominence, pitch range, environment, and layering using the method and scales outlined above, I can precisely compare how both artists' voices unfold in virtual space both within and across different songs. While I focus exclusively on the first verse and chorus of each song, the same method could be applied to entire tracks.

[4.3] Eminem's and Rihanna's four collaborations are especially suitable objects of study for comparing vocal placement between two artists. First, the four songs were released within a span of three years, ensuring that there is stylistic continuity between each track. Second, each artist appears twice as a main artist and twice as a guest artist. This equal division of musical labor is reflected in the songs' formal structures (**Example 16**). When performing as an invited guest, Rihanna provides the chorus; Eminem's role as a featured artist is to rap the third verse.[36] This parallelism between songs can allow us to determine whether Rihanna's voice, for instance, is spatialized differently depending on her status as a main or a guest artist. Finally, the two versions of "Love the Way You Lie" are especially suited for comparison as they engage in direct dialogue with one another. The songs share the same chorus, and are conceived as mirror images of one another, with Eminem and Rihanna acting respectively as narrators.[37]

Width

[4.4] **Example 17** displays the vocal placement of Rihanna and Eminem in the first verse and chorus of their four collaborations. **Example 18** summarizes the data from Example 17. To compare vocal placement in each song, I discuss each sonic parameter separately. In the four tracks, Eminem's and Rihanna's voices are consistently centered within the virtual space. As acknowledged by Moylan (2015) and Zak (2001), this placement is conventional for recorded popular music. Focusing on the width of both artists' voices reveals an interesting trend: Rihanna's voice is always presented as significantly more diffuse than Eminem's, with width ratings of W3, W4, and W5.

Eminem's voice, conversely, always occupies a narrower position than Rihanna in the virtual space. Despite minute variations in the width of his voice from song to song, his voice consistently sounds more localized with width ratings of W1, W2, and W3. Rihanna's voice is presented in more reverberant spaces, which accounts for its more diffuse presence within the virtual space. Her singing voice is centered, but the delayed images of her voice are panned toward the sides of the stereo image.

Pitch Range

[4.5] Rihanna's voice unfolds within a G3 to D5 pitch range across the four tracks. Eminem's rapping voice covers an approximate range of C♯3 to F♯4.

Prominence

[4.6] In all songs, apart from "Numb," Rihanna's voice is more salient than Eminem's. Her voice often appears nearly alone in the virtual space, only supported by a low-amplitude piano or synthesizer. Her voice therefore takes on a prominent role in comparison to the other sound sources heard in the mix.[38] Eminem's voice, on the contrary, is often accompanied by a thicker instrumental texture. His voice is therefore more blended within the virtual space.

Environment

[4.7] Rihanna's voice is always presented in a more reverberant environment than Eminem's. The distinction between their voices is made especially evident in "The Monster" and "Love the Way You Lie," in which Rihanna's reverberant vocal environment E5 contrasts with Eminem's flat environment of E1.

Layering

[4.8] There are few instances of clear and persistent layering in the four songs under study. In "Numb," however, Rihanna's voice is heavily layered (L5) with harmonies. In "Love the Way You Lie (Part II)" and "Numb," Eminem's voice is subtly ornamented with added layers.

[4.9] The brief analytical comparisons outlined above precisely describe how the vocal placements of Rihanna and Eminem differ throughout their four collaborations. Rihanna's voice tends to be wider, higher, and more salient that Eminem's; her voice also tends to be placed in more reverberant environments and ornamented with more layers. The stark contrast between Rihanna's and Eminem's vocal placements in virtual space raises the following questions: why might Rihanna's and Eminem's voices evoke such different spatial characteristics? What is the effect of these contrasting vocal placements?

[4.10] The differences in vocal placement may be attributed to mixing conventions that aim to maximize vocal contrast in collaborations between a rapper and a singer. Each track has been recorded, mixed, and produced at a variety of studios, by different personnel (**Example 19**), lowering the possibility that vocal placement is based on a single engineer's idiosyncratic artistic choices rather than on larger-scale conventions regarding mixing norms within U.S. popular music. For instance, the rapped verses for "Love the Way You Lie" were recorded at Eminem's own Effigy Studios. Rihanna recorded her chorus while on tour, in Dublin's Sun Studios. Mixing engineer Mike Strange explains that he "kept the vocal balance made by Rihanna's people, and simply had Rihanna's vocals come up on the desk in stereo" (Herrera 2010). Elsewhere (Duguay 2021), I show that most collaborations between a woman who sings and a man who raps follow the same type of vocal placement configuration as the ones found in the Rihanna/Eminem collaborations: a wide and reverberated woman's voice alongside a narrow and flat man's voice. I hear this configuration as a constructed sonic contrast that presents women's voices as ornamental and diffuse, and men's voices as direct and relatable. These contrasting vocal placements perhaps contribute to the sonic construction of a gender binary, exemplifying one of the ways in which dichotomous conceptions of gender are reinforced through the sound of popular music.

[4.11] This article has outlined a methodology for precisely locating a voice's placement within the virtual space created in a recording. Analysts can use the methodology to create clear visual representations of a voice's width, pitch range, prominence, environment, and layering. The approach may be of use to any analyst wishing to precisely describe vocal placements in recordings. Because each parameter is measured according to a pre-defined scale, the methodology can also be used for comparing two or more vocal placements within the same recording or across different tracks. Finally, while the methodology currently focuses on vocal placement, it could be easily adapted to study the placement of any recorded sound source as it unfolds within virtual space.

[4.12] As a means of conclusion, I identify four ways in which the methodology outlined in this article could be further developed and refined. First, it could be refined to precisely reflect the temporal aspect of a voice's spatial organization. The visualizations I provide in this article, as static images, show a cumulative representation of a voice's vocal placement as it evolves over a span of time. Video animations of vocal placement, however, could more accurately show minute variations in the spatialization of voices as they unfold through time. Second, the methodology does not explicitly consider the interactions between the voice and other sound sources in the recordings.[39] As a solution, one could use the methodology to analyze the placement of additional sound sources within the virtual space. Precise discussions of the ways in which sound sources interact with one another would allow for a more refined understanding of the auditory impact of different vocal placements. A particular vocal placement, for instance, may have a different impact when placed within a full or empty virtual space. Third, the methodology currently does not consider the size of the virtual space. To enable comparisons between tracks, I show vocal placements as they fit within a standardized "sound-box." The apparent dimensions of this sound-box, however, may seem to vary both within and across tracks. Two vocal placements could *look* identical, with the same pitch range, environment, prominence, layering, and width data, but *sound* different. One of these performances may be whispered and closely miked, therefore emulating a small, enclosed space; the other may be almost shouted and reminiscent of a live performance on a public stage. An analyst could account for these discrepancies by considering aspects of vocal performance (Heidemann 2016; Malawey 2020; Moore, Schmidt, and Dockwray 2011) alongside with vocal placement.

[4.13] Finally, the methodology can be time consuming because it relies on several different analytical tools. This potential difficulty could be circumvented by the development of a program that automates the audio feature extraction components of the methodology. With an automation of this portion of the process, an analyst could then obtain a visualization of a voice's placement after answering a few questions based on aural analysis. I am confident that future work in the study of vocal placement and virtual space will make this analytical method even more accessible to music researchers; as of now, I hope to have provided a consistent and reliable way to engage with one of the most crucial parameters of recorded popular music.

Michèle Duguay
Indiana University
Simon Music Center, M225E
1201 E 3rd St
Bloomington, IN 47405
mduguay@iu.edu

Works Cited

Brøvig-Hanssen, Ragnild, and Anne Danielsen. 2013. "The Naturalized and the Surreal: Changes in the Perception of Popular Music Sound." *Organized* Sound 18 (1): 71–80. https://doi.org/10.1017/S1355771812000258.

—————. 2016. *Digital Signatures: The Impact of Digitization on Popular Music Sound*. MIT Press. https://doi.org/10.7551/mitpress/10192.001.0001.

Burns, Lori, and Alyssa Woods. 2004. "Authenticity, Appropriation, Signification: Tori Amos on Gender, Race, and Violence in Covers of Billie Holiday and Eminem." *Music Theory Online* 10 (2). https://www.mtosmt.org/issues/mto.04.10.2/mto.04.10.2.burns_woods.html.

Camilleri, Lelio. 2010. "Shaping Sounds, Shaping Spaces." *Popular Music* 29 (2): 199–211. https://doi.org/10.1017/S0261143010000036.

Cannam, Chris, Mark Sandler, Michael O. Jewell, Christophe Rhodes, and Mark d'Inverno. 2010. "Linked Data and You: Bringing Music Research Software into the Semantic Web." *Journal of New Music Research* 39 (4): 313–25. https://doi.org/10.1080/09298215.2010.522715.

Clarke, Eric. 2013. "Music, Space and Subjectivity." In *Music, Sound and Space Transformations of Public and Private Experience*, ed. Georgina Born, 90–110. Cambridge University Press. https://doi.org/10.1017/CBO9780511675850.004.

Danielsen, Anne. 1993."My name is Prince": *En studie i Diamonds and Pearls* [*A study of Diamonds and Pearls*]. Second degree level thesis, University of Oslo.

—————. 1997. "His Name Was Prince: A Study of Diamonds and Pearls." *Popular Music* 16 (3): 275–91. https://doi.org/10.1017/S0261143000008412.

Dibben, Nicola. 2013. "The Intimate Singing Voice: Auditory Spatial Perception and Emotion in Pop Recordings." In *Electrified Voices: Medial, Socio-Historical and Cultural Aspects of Voice Transfer*, ed. Dmitri Zakharine and Nils Meise, 107–22. V&R Unipress.

Doyle, Peter. 2005. *Echo and Reverb: Fabricating Space in Popular Music Recording, 1900–1960*. Wesleyan University Press.

Duguay, Michèle. 2021. "Gendering the Virtual Space: Sonic Femininities and Masculinities in Contemporary Top 40 Music." PhD diss., The Graduate Center, CUNY.

Dockwray, Ruth, and Allan F. Moore. 2010. "Configuring the Sound-Box." *Popular Music* 29 (2): 181–97. https://doi.org/10.1017/S0261143010000024.

Gamble, Steve. 2019. "Listening to Virtual Space in Recorded Popular Music." In *Proceedings of the 12th Art of Record Production Conference Mono: Stereo: Multi*, ed. Jan-Olof Gullö, Shara Rambarran, and Katia Isakoff, 105–18. Royal College of Music (KMH) & Art of Record Production.

Gibson, David. 2005. *The Art of Mixing: A Visual Guide to Recording, Engineering, and Production*. Thomson Course Technology.

Greitzer, Mary Lee. 2013. "Queer Responses to Sexual Trauma: The Voices of Tori Amos's 'Me and a Gun' and Lydia Lunch's *Daddy Dearest*." *Women and Music: A Journal of Gender and Culture* 17: 1–26. https://doi.org/10.1353/wam.2013.0000.

Hall, Edward T. 1963. "A System for the Notation of Proxemic Behavior." *American Anthropologist* 65 (5): 1003–26. https://doi.org/10.1525/aa.1963.65.5.02a00020.

Heidemann, Kate. 2014. "Hearing Women's Voices in Popular Song: Analyzing Sound and Identity in Country and Soul." PhD diss., Columbia University.

—————. 2016. "A System for Describing Vocal Timbre in Popular Song." *Music Theory Online* 22 (1). https://mtosmt.org/issues/mto.16.22.1/mto.16.22.1.heidemann.html.

Herrera, Monica. 2010. "Hitmaker Alex Da Kid Talks Eminem's 'Lie,' Rihanna's New Album." *Billboard*, August 8. https://www.billboard.com/music/music-news/hitmaker-alex-da-kid-talks-eminems-lie-rihannas-new-album-956815/.

Huber, David Miles. 2018. *Modern Recording Techniques*, 9th ed. Routledge.

Irwing, Paul, and David J. Hughes. 2018. "Test Development." In *The Wiley Handbook of Psychometric Testing*, ed. Paul Irwing, Tom Booth, and David J. Hughes, 3–48. Wiley Blackwell. https://doi.org/10.1002/9781118489772.

Kraugerud, Emil. 2020. "Come Closer: Acousmatic Intimacy in Popular Music Sound." PhD diss., University of Oslo.

Lacasse, Serge. 2000. "'Listen to My Voice': The Evocative Power of Vocal Staging in Recorded Rock Music and Other Forms of Vocal Expression." PhD diss., University of Liverpool.

Lacasse, Serge, and Tara Mimnagh. 2005. "Quand Amos se fait Eminem: Féminisation, intertextualité et mise en scène phonographique." In *Le féminin, le masculin et la musique aujourd'hui: Actes de la journée du 4 mars 2003*, ed. Cécile Prévost-Thomas, Hyacinthe Ravet, and Catherine Rudent, 109–17. Observatoire musical français.

Malawey, Victoria. 2020. *A Blaze of Light in Every Word: Analyzing the Popular Singing Voice*. Oxford University Press. https://doi.org/10.1093/oso/9780190052201.001.0001.

Mauch, Matthias, and Simon Dixon. 2014. "PYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions." In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), 659–63. https://doi.org/10.1109/ICASSP.2014.6853678.

McNally, Kirk, George Tzanetakis, and Steven R. Ness. 2009. "New Tools for Use in the Musicology of Record Production." Unpublished Paper, University of Victoria. https://www.academia.edu/23107006/New_tools_for_use_in_the_musicology_of_record_production?source=swp_share.

Moore, Allan F. 1993. *Rock: The Primary Text*. Open University Press.

Moore, Allan F., Patricia Schmidt, and Ruth Dockwray. 2011. "A Hermeneutics of Spatialization for Recorded Song." *Twentieth-Century Music* 6 (1): 83–114. https://doi.org/10.1017/S1478572210000071.

Moylan, William. 2009. "Considering Space in Music." *Journal on the Art of Record Production* 4. https://www.arpjournal.com/asarpwp/considering-space-in-music/. https://doi.org/10.4324/9780203758410.

—————. 2015. *Understanding and Crafting the Mix: The Art of Recording*. Focal Press.

Owsinski, Bobby. 2017. *The Mixing Engineer's Handbook*. Thomson Course Technology.

Sofer, Danielle. 2020. "Specters of Sex Tracing the Tools and Techniques of Contemporary Music Analysis." *Zeitschrift der Gesellschaft für Musiktheorie* 17 (1): 31–63. https://doi.org/10.31751/1029.

Stöter, Fabian-Robert, Stefan Uhlich, Antoine Liutkus, and Yuki Mitsufuji. 2019. "Open-Unmix: A Reference Implementation for Music Source Separation." *Journal of Open Source Software* 4 (41). https://doi.org/10.21105/joss.01667.

Tillet, Salamishah. 2014. "Strange Sampling: Nina Simone and Her Hip-Hop Children." *American Quarterly* 66 (1): 119–37. https://doi.org/10.1353/aq.2014.0006.

Zagorski-Thomas, Simon. 2014. *The Musicology of Record Production*. Cambridge University Press. https://doi.org/10.1017/CBO9781139871846.

Zak, Albin. 2001. *The Poetics of Rock: Cutting Tracks, Making Records*. University of California Press. https://doi.org/10.1525/9780520928152.

Zinser, Zachary. 2020. "Sound, Syntax, and Space in Studio-Produced Popular Music." PhD diss., Indiana University.

Footnotes

1. Released in 2013 on *The Marshall Mathers LP 2*.
Return to text

2. The definition of vocal placement offered here is distinct from the use of "vocal placement" in vocal pedagogy and technique. In this context, vocal placement refers to the physiological space controlled by the singer to change the color and timbre of their voice.
Return to text

3. This review of the existing work on virtual space focuses on analytical literature that takes the finished recording as its point of departure. Studies on the process of recording and mixing also address the formation of virtual space (Moylan 2009 and 2015), but I do not address them here in detail. Additionally, I choose to focus my discussion on studies of virtual space that propose an original analytical methodology.
Return to text

4. I use the term virtual space throughout the article to refer to the abstract space evoked by a recording. Not all authors discussed here, however, use the same term to refer to this phenomenon. Dockwray and Moore (2010) and Danielsen (1997), for instance, speak of the "sound-box," Lacasse of "Vocal Staging," and Malawey of "Mediation with Technology."
Return to text

5. Dockwray and Moore use the sound-box to track the development of stereo placement in commercial recordings. They provide a taxonomy of mixes used in pop-rock, easy listening, and psychedelic tracks from 1966 to 1972. Their study highlights how the "diagonal mix"—in which the vocals, bass, and snare drum are positioned in a diagonal configuration in relationship to the vertical axis—became normalized in the 1970s.
Return to text

6. Kraugerud (2020, 22) notes that Danielsen's concept was originally called *lydrom* ("sound room") (1993), and later translated to sound-box (1997).
Return to text

7. For another comparison of Danielsen's and Moore's methods, see Kraugerud 2020, 22–23.
Return to text

8. In French-language literature, Lacasse refers to this concept as "la mise en scène phonographique" (Lacasse and Mimnagh 2005).
Return to text

9. Since Zinser explicitly states that his analyses "do not reflect any attempt at constructing a rigorous methodology" for analyzing virtual space, I do not discuss his analyses in detail here (2020, 41). His dissertation is nonetheless a milestone in the study of virtual space in recent popular music, replete with various visual representations that evocatively capture the recorded space as it is experienced by a listener.
Return to text

10. The four zones are determined by (1) the perceived distance between performer and listener; (2) the location of the voice in the sound environment; and (3) the articulation of the singing persona based on lyrics and vocal delivery.
Return to text

11. Nicola Dibben (2013) builds on Moore, Schmidt, and Dockwray's four zones to examine the relationship between listener and singer. The latter is often centered in the mix, and Dibben attributes this norm to a music industry predicated on celebrity. Ultimately, she concludes that the aural proximity afforded by intimate virtual spaces invites a focus on the private lives of the artist.
Return to text

12. Example 5 identifies the parameters named by the authors as explicitly linked to virtual space. The author(s) may address other parameters in their discussions of virtual space, but I decided to

include in Example 5 only the ones that are explicitly described as central to their methodology.
Return to text

13. Camilleri acknowledges that the impression of depth relates to "the spectral content (timbre) of sound, [which] plays a relevant role in the overall perception of space" and "the combination of spectral content of sounds and their disposition" (Camilleri 2010, 202). His analyses, however, focus on depth as reverberation.
Return to text

14. Moore, Schmidt, and Dockwray's 2009 classification of potential distances between the singer and listener is a useful step toward such a scale. They identify four proxemic zones, after Hall (1963), in which a piece of recorded music can unfold: (1) intimate (very close to the listener); (2) personal (close to the listener); (3) social (medium distance from the listener); and (4) public (sizeable distance from the listener). Upon listening to a track, an analyst can instinctively tell if the singer is located at a close or sizeable distance. A detailed comparison of two tracks, however, would require a precise way to measure distance.
Return to text

15. Alternatively, one could conceive of a situation in which the listener mentally positions themselves as one of the sound sources in the virtual space, such as the main vocals. In this case, the listener may experience the recorded performance as though they were on the virtual stage.
Return to text

16. While the analyses presented in this article were conducted using headphones, an analyst could also use (stereo) loudspeakers to listen for vocal placement. This analyst would need to be situated in the "sweet spot," or the focal point between the two speakers that allows for the stereo mix to be fully heard. While headphones and loudspeakers create different listening experiences, the differences would not significantly alter the analysis of vocal placement within virtual space. The methodology assumes a somewhat idealized listening position, in which the listener has access to a high-quality sound reproduction technology. Various parameters, such as playback technology, listening environment, and hearing loss may affect an analyst's perception of the virtual space. More research would be needed to study the ways in which varying listening positionalities shape virtual spaces.
Return to text

17. Likert and Likert-type scales are rating scales frequently used in questionnaires and surveys. Participants are asked to specify their thoughts on a given issue according to a 5- or 7-point scale (strongly disagree, disagree, neutral, agree, strongly agree are frequently used categories). Likert-type scales, like the ones used for measuring environment, width, and layering, are used to answer a single question. A full Likert scale considers the sum or average of responses over a set of individual questions (Irwing 2018, 9).
Return to text

18. Sonic Visualiser is a free, open-source software for the visualization and analysis of audio recordings. It was developed in 2010 by Chris Cannam, Christian Landone and Mark Sandler at Queen Mary University of London's Centre for Digital Music (Cannam et al. 2010). Users can import, edit, and annotate audio files on its Desktop user interface, which also supports a variety of plug-ins for sound processing. See https://www.vamp-plugins.org/download.html for a list of available plug-ins.
Return to text

19. The source code for Open-Unmix can be downloaded and used on Linux or OSX, and the website also provides a tutorial for installation and usage.
Return to text

20. Commercial popular music recording practice tends to place lead vocals at the center of the virtual space. This central panning ensures a "greater psychoacoustic impact" in which the voice is the most salient aspect of a recording (Zak 2001, 155). Since voices in recent popular music are

positioned to the center of the virtual space, the lateral position (panning) of voices will not be discussed at length as a standalone parameter throughout this article. The parameter of width, while related to panning, captures instead the breadth of a voice as it appears on the stereo stage.
Return to text

21. I equate the vertical axis with pitch for three reasons: (1) Equating the vertical axis of the virtual space to pitch invokes traditional metaphors of pitch organization in Western music notation, in which frequencies are organized on a vertical scale from lowest to highest. Dockwray and Moore's (2010) and Danielsen's (1997) approaches to the sound-box follow suit, along with several sound engineering manuals (Gibson 2005; Huber 2018; Owsinski 2017); (2) While humans can discriminate sound based on its horizontal location, the height—in actual location, not pitch—of a sound is more difficult to perceive. Since humans can discriminate between pitch height much better, it follows that the virtual space's vertical axis should represent pitch; (3) The third reason is mainly practical: without a vertical axis showing pitch, the virtual space would be flattened to two dimensions and therefore more difficult to represent visually.
Return to text

22. The PYIN algorithm, developed by Matthias Mauch and Simon Dixon, tracks the estimated frequency of a sound signal over time (2014). When used as a plug-in in Sonic Visualiser, the "PYIN Smoothed Pitch Track" function outputs an estimation of a melody's pitches. Note that the plug-in is meant to analyze monophonic vocal tracks and can only capture one pitch per time frame. When a voice is layered with harmonies, "Smoothed Pitch Track" will output what it estimates to be the main melodic line.
Return to text

23. Amplitude is a measurement of the size (height) of a sound wave. Amplitude is related to, but not analogous to volume. If the amplitude increases, a sound is perceived as becoming louder; if the amplitude decreases, a sound is perceived as becoming softer.
Return to text

24. The RMS Amplitude plug-in is part of the "LibXtract Vamp plugins" group, which implements several functions of Jamie Bullock's LibXtract library. The library contains various audio feature extraction functions. The plug-in can be accessed here https://code.soundsoftware.ac.uk/projects/vamp-libxtract-plugins, and the LibXtract library can be accessed here https://github.com/jamiebullock/LibXtract. The RMS Amplitude (root mean square amplitude) function computes the amplitude of a sound signal over time. The function first establishes the square root of the arithmetic mean of the sound signal, then squares the result. This process ensures that the amplitude of a sound signal—which has both positive and negative values—is expressed as a set of positive values only.
Return to text

25. This value was obtained from the isolated vocal track and consists in the average of all the RMS amplitude values over time in the chorus. RMS amplitude values lower than 0.02 were omitted in the calculations of the average, to avoid considering moments of silence or near-silence in the vocal track.
Return to text

26. This value was obtained from the full track and consists in the average of all the RMS amplitude values over time in the chorus.
Return to text

27. Camilleri, for instance, conceptualizes it in terms of reverberation, Dockwray and Moore in terms of volume and reverberation, and Lacasse through another combination of parameters. The depth axis is also discussed differently in several sound engineering manuals. In the *Mixing Engineer's Handbook*, it is described as the "depth" that can be achieved through reverberation and delay (Owsinski 2017, 156). *The Art of Mixing* shows volume as the responsible parameter for front-back location. Gibson nonetheless admits that the ear "need(s) other cues, such as delays and reverb, to help gauge the distance [of a sound] (Gibson 2005, 24). *Modern Recording Techniques*

center the importance of reverberation, which can be used by the engineer to "place the sound source at either the front or the rear of the artificially created soundscape" (Huber 2018, 74).
Return to text

28. Moylan (2015) describes the parameters that are not responsible for the perception of distance in recorded music: (1) loudness, (2) the amount of reverberation, and (3) microphone technique. First, while a sound loses amplitude as distance increases, loudness does not directly relate to distance. A soft sound may be close, and a loud sound may be far. Second, the amount of reverberation does not contribute to distance perception. A reverberant sound may be close while a dry sound may be far. As discussed above, it is rather the quality of the reverberation—the amplitude of the direct sound in relationship to the reverberated images—that contributes to the perception of distance. Finally, the perceived distance between a microphone and a sound source does not contribute to the perception of distance; microphone technique only accounts for the level of timbral detail captured.
Return to text

29. It would be possible to study a mixing engineer's manipulation of distance in a recording, but only with access to the DAW files in addition to the completed mix. One could then compare the original, unprocessed recording to the final product.
Return to text

30. Reverberation differs from echo in that the reflections occur before approximately 50ms. Echoed repetitions of a sound are easily detectable by ear, with a minimum reflection time of 50ms. Brøvig-Hanssen and Danielsen provide an evocative description of such temporal effects, noting that "in an enclosed physical space, such as a room, sound travels in all dimensions and bounces around as it meets the surfaces of walls, the floor, or the ceiling in turn. The multiple (and multiplying) sonic reflections of the sound gradually weaken as the air and the surfaces absorb them, until they die out entirely. If a sound hits its first surface (preferably a hard one) after fifty to eighty milliseconds (depending on the sound itself), its reflections will be audible as distinct, separated sounds—what we refer to as echo or delay . . . Enclosed spaces that produce echo might include large empty buildings or wells; outside spaces might include neighboring mountains or tall concrete walls. More common than the echo effect (in nonmusical circumstances) is, of course, the effect of reverb — that is, when a sound hits a surface immediately and promptly hits many others as well (a ping-pong effect), producing multiple echoes that are so dense that the reflections cannot be distinguished from one other" (2016, 23). For another in-depth description and comparison of echo and reverberation, see Lacasse (2000).
Return to text

31. The aspects of reverberation captured by this recording would depend on the microphone's type and position.
Return to text

32. The excerpts discussed in Examples 12, 16, and 17, along with the four collaborations between Rihanna and Eminem, are part of my Collaborative Song (CS) Corpus (Duguay 2021). The CS corpus contains all musical collaborations that appeared on the 2008–18 *Billboard* Year-end Hot 100 charts and include at least two vocalists of different genders.
Return to text

33. I would like to thank Lizhou Wang for his astute observation that echoes such as the ones found in E5 can sometimes act as a form of vocal layering. When an echo is not treated with reverberation, the dry repetition of the main vocal line is perhaps best described as an added vocal layer than as a spatial manipulation of the recorded voice. Throughout the article, I nonetheless choose to categorize such an effect as the E5 category because I believe that echo and reverberation, while not completely analogous, are temporal manipulations of a sound that are best considered together. An analyst may nevertheless choose to ignore the E5 category and treat such echo effects as a form of layering.
Return to text

34. A recorded voice is often superposed with one or more identical copy of itself, in a practice known as double-tracking. In many cases, double-tracking is used in the mixing process to create a fuller sound, with a final result sounding like a single vocal layer. I only consider double-tracking in my analysis if the vocal layers are obvious to the ear. Additionally, the analyses of layering outlined in this article only consider copies of the singer's own voice and does not take into consideration added background vocals provided by one or more additional singers. An analyst may choose, however, to incorporate background vocals to their own analyses.
Return to text

35. In the verse of "We Found Love" (2011), for instance, Rihanna's voice has a single layer. While highly reverberated, her singing voice is not layered by other vocal tracks, such as in 0:08–0:38. In some instances, the layering can be especially subtle. Cardi B's voice in "I Like It" (2018) might at first appear to stand alone, but careful listening reveals the existence of other vocal tracks layered at a lower amplitude, as heard in 0:14–0:26.
Return to text

36. Both formal outlines are typical for songs featuring a rapper and a singer.
Return to text

37. Elsewhere, I examine the two versions of "Love The Way You Lie" with a focus on the narratives of domestic violence that arise from the lyrics, vocal delivery, and vocal placement heard in both songs (Duguay 2021).
Return to text

38. In subsequent choruses, her voice takes on a lower prominence as other accompanying instruments join the mix.
Return to text

39. The prominence parameter is an exception, because it measures the voice's amplitude in relationship to that of the other sound sources in the mix.
Return to text