Immersed in Pop!

Excursions into Compositional Design

Zack Bresler

Zack Bresler

Immersed in Pop!

Excursions into Compositional Design

Dissertation for the degree Philosophiae Doctor

University of Agder

Faculty of Fine Arts

2021

Acknowledgements

Thank you first to the faculty and administration at the Department of Popular Music, University of Agder, for providing me with the time, resources, and scholarship to carry out this study. It has been a tremendous privilege and honor to be given such an opportunity.

From the bottom of my heart, thank you to my principal supervisor Professor Stan Hawkins. Your relentless support has inspired me on numerous occasions, and your wit and passion has been contagious. I feel proud, honored, and truly privileged to have worked with you, both as your student and as co-author, and I am happy to call you not only my colleague, but also my friend.

Thank you to my co-supervisor Jon Marius Aareskjold, who was particularly helpful in the development of my practical knowledge around immersive and interactive media. It has been awesome to work with you on installations and performances of immersive music, and I hope now that I will have more time for these projects.

To my colleagues at the Department of Popular Music, University of Agder – I feel fortunate to have produced this work alongside you. Thanks for all your conversations–in the hallways, in the lunchroom, and in the seminar, especially those of you in my PhD cohort, Andreas, Kari, Vincent, Eirik, Gunn-Hilde, and Bodil. You are all part of this thesis, whether you know it or not, and I wish you all the best of luck in your futures.

Also, thanks to my friends and colleagues at the numerous academic organizations and conferences that I have had the fortune to be a member of or present at, including the Audio Engineering Society, the Art of Record Production, and IASPM (and in particular the Nordic branch).

Finally, I want to say thank you to my wife and best friend, Maggie. You inspire me every day with your support, care, and love. Thanks for everything, not least listening to my presentations, proofreading my articles, critiquing my slideshows, discussing my ideas, hearing my frustrations, supporting me through challenges, and celebrating my wins (both big and small). I couldn't have done it without you. Love you, boo.

Summary

English:

Recent changes in consumer audio and music technology and distribution—for example the addition of 3D audio formats such as Dolby Atmos to music streaming services, the recent release of "Spatial Audio" on Apple and Beats products, the proliferation of musical content in virtual reality and 360° videos, etc.—have reignited a public discourse on concepts of immersion and interactivity in popular music and media. This raises questions and necessitates a deepening of popular musicological discourse in these areas. This thesis thus asks: what is the relationship between so-called immersive media and immersive experience? How are immersive and interactive experiences of audiovisual popular music compositionally designed? And to what degree do interpretations of immersion and interactivity in popular music imply agency on part of the listener/viewer? To address these questions, Bresler has authored or co-authored four articles and book chapters on music in immersive and interactive media with a focus on compositional design and immersion in pop music. In the framing chapter, these articles are contextualized through the coining of the term *immersive staging*, which is a framework for understanding how the perceived relationship between the performer and listener is mediated through technology, performativity, audiovisual compositional design, and aesthetics. Additionally, the chapter makes a case for the hermeneutic methodologies employed throughout.

Norsk:

Nylig utvikling innen forbrukerlyd, musikkteknologi og distribusjon – for eksempel tillegg av 3D lydformater som Dolby Atmos til streaming tjenester, den siste utgaven at «Spatial Audio» på Apple og Beats -produkter, spredning av musikkinnhold i virtual reality og 360-videoer, etc. – her skapt en offentlig diskurs om konsepter rundt *immersjon* og interaktivitet i popmusikk og media. Dette stiller essensielle spørsmål, og nødvendiggjør samtidig utviklingen av en musikologisk diskurs på temaene. Denne avhandling spør derfor: hva er forholdet mellom såkalte immersive medier og immersive opplevelse? Hvordan er immersive og interaktive opplevelser av audiovisuell popmusikk designet? Og til sist, i hvilken grad innebærer tolkninger av immersjon og interaktivitet i popmusikk agens (*agency*) for lytteren/seeren? For å løse disse spørsmålene har Bresler forfattet og medforfattet fire artikler og bokkapitaler om musikk i immersive og interaktiv

medier med fokus på komposisjonell design og immersjon i popmusikk. I kappen blir disse artiklene kontekstualisert gjennom begrepet *immersive staging*, som er et rammeverk for å forstå hvordan det oppfattede, perseptuelle forholdet mellom utøveren og lytteren formidles gjennom teknologi, performativitet, audiovisuell komposisjonell design, og estetikk. I tillegg argumenterer innholdet i kappen for de hermeneutiske metodene som brukes gjennomgående.

Contents

Introduction

It is late 1973, and a hi-fi enthusiast has just gotten their hands on Pink Floyd's *Dark Side of the Moon*. "*Finally,*" she thinks, having already purchased the stereo vinyl earlier this year, "*I cannot wait to hear this in quad.*" Making sure the SQ button is depressed on the decoder, she places the record immediately on Side B, eager to hear her favorite track 'Money' in the new format. In stereo, the coins and cash register sounds that open the track pan wildly with each new sound of the 7/4 loop changing direction quickly and confounding the normal sense of space in rock recordings. In quad, the track is even more exciting, as each of the sounds seems to emanate discretely from its own speaker, replete with its spatial characteristics. The result is that the listener is truly surround by a cacophony of capitalism—the sounds of coins dropping and cash registers ringing out in all directions until the guitar begins with its famous riff in the rear left speaker. After a couple repeats, the drums hit a single lead-in on beat 7, cueing the full band entrance. Now, the drums are panned in a kind of stereo in front with the bass guitar centered, the dimensions of the stage are finally established. "*Cool.*"

Pink Floyd's *Dark Side of the Moon* was one of many records released in the 1970s on the quadraphonic surround format. While quadraphonic ultimately failed,[1] the record was particularly innovative, having been engineered by Alan Parsons with quad specifically in mind. It's having been conceived this way is well documented by Parsons himself, but also evidenced in the band's reaction to the album being released initially in stereo only on a rushed schedule, since they chose not to show up to the release party (Povey, 2016, p. 210). At any rate, although the final mixes on the record were credited to Chris Thomas, Parsons was instrumental in the conception of its quadraphonic construction and his strategies for the recording and production in quad have been well documented. Writing about it himself in 1975, he said that although the record "was monitored in studios equipped for stereo reproduction, many sections were recorded with regard to the eventual quadraphonic" (1975). And in an interview in 2002, he reiterated his

---

[1] Quad's failure could be seen as a foreshadowing, as stereo has dominated popular music since it took hold and subsequent attempts at bringing multi-channel into mainstream popular music (quad, 5.1 surround, binaural, etc.) have to this point failed, at least in a commercial sense. However, this is beside the point here, since this music was certainly successful artistically and continues to be influential.

opinion that "the surround experience shouldn't be a stereo experience with ambience. It should be four stereo sound fields… I liked the idea of action happening in all four channels. I wasn't particularly interested in it sounding like a band onstage" (Parsons, quoted in K. Richardson, 2003).

This example of *Dark Side of the Moon* in quadraphonic demonstrates many characteristics of the themes that underpin this thesis. For one, it shows that, although current discourse often suggests a progressive narrative about musical spatiality, innovation and experimentation in multichannel audio formats is not new. Decades after its initial release, the surround and quad mixes of Dark Side were re-released on a Blu-Ray box set in 2011, and Parson's mixes were still reviewed in the press as being innovative and captivating.[2] The debates among artists, producers, and mixers on how to best use space in mixes have been always ongoing and mediated through technological innovations. Rather than being a fixed entity with long-ago settled norms, spatial audio in every format from stereo to Dolby Atmos is always changing, not only with the tastes of artists and listeners, but also through the possibilities afforded by the technologies that enable pop music composition, production, and dissemination.

Second, it demonstrates the way that immersive music (a term which I will define later) can dramatically alter the way listeners and artists are engaged. While stereophonic technology certainly opened the sonic space from mono into three dimensions, surround sound and later 3D audio has allowed for the literal envelopment of the listener with sound. While the possibilities this affords recordists are boundless, so too are the ways this implies new modes of reception, compositional design, performativity, and staging. Considering the listener's experience as described above, it is clear that the changing presentation of the sonic material can have great implications for the way the listener identifies with the music, from the perceived proximity to sound sources and their new positioning in space to the meanings of lyrics and melodies as they shift around and through the listener.

Finally, the example highlights the importance of subjectivity and intertextuality in the interpretation of the pop score. Every aspect of the listening situation, including temporal specifics (Hawkins, 2016, p. 2) of the listener, her gendered/racial/class identity, the importance of the music within the popular

---

[2] https://theseconddisc.com/2011/10/06/review-pink-floyd-the-dark-side-of-the-moon-immersion-box-set/

culture, the listener's intended function for the music, even their mood and affective state, all play a role in musical interpretation. In the example above, our hypothetical listener is not only clearly a fan of the music, but also of the technology surrounding quad—a niche and expensive hobby in 1973. She is listening to the music as the primary activity, rather than using it as a device for something else like reading a book or sharing a meal. Moreover, her listening is informed by intertext, having already heard the stereo version (presumably both on her home speakers and in other contexts such as the radio or on television) and comparing it for effect.

The framing chapter makes the argument for considering immersion in audiovisual pop music, working within what I am calling a musicology of immersion and coining the term immersive staging, which is framework for understanding how the perceived relationship between the performer and listener is mediated through technology, performativity, audiovisual compositional design, and aesthetics. Additionally, it attempts to unite the accompanying articles into one project and makes a case for the hermeneutic methodologies employed throughout.

This subject is deeply meaningful to me. I have been engrossed in music and technology ever since I can remember. I latched on to drumming at a young age, and since I first sat down at a computer I have been drawn into the mysteries and secrets of digital technology. In my adolescence I became deeply interested in music recording, digital audio workstations, microphones, MIDI, synthesis, and electronic instruments. In university I became a classically trained percussionist, primarily because I knew of no other path through academia, and it was not until I was in my master's in music performance at the University of Nebraska at Omaha where I studied under Scott Shinbara that I realized that my love of music and technology could be bridged in an academic environment. It was Scott who first encouraged me to experiment with technology in performance, for example playing pieces for percussion and tape, live digital signal processing, and MIDI triggering.

Among the many things I learned in those years is that a love for popular music could be acceptable for a music academic. Scott constantly referenced pop music and culture in our private lessons (like the time he told me to channel Miley Cyrus at the marimba, because "when we are performing, we can't stop and we won't stop"), and he regularly and openly challenged problematic and elitist ideas on the superiority of western classical music and music theory. This was something that

I had not seen in academia to that point, excepting the occasional reference to artists like Frank Zappa or The Beatles (and in those cases the implication was often that they were 'ok' to be into because of their adherence to the virtues of western tonality). He introduced digital performances of pop music albums on MIDI percussion controllers as a regular features of the UNO Percussion Ensemble, and I will never forget our performances of Radiohead's In Rainbows, which we did in collaboration with local pop music singers in 2013.

It was, in a huge way, these experiences from my years studying at UNO that led me first to teaching music production and recording, then to come to Norway for further study where I would ultimately write this thesis in what is I hope an interesting and important contribution in popular musicology. While this journey has been an unexpected career path for me (as seems to be the case with many who have entered this field), I ultimately feel lucky to have fallen into the discipline of popular musicology, which combines all my interests so deeply—music technology, pop music and culture, music production aesthetics, and most of all the audiovisual spectacle of pop music performances.

Research Questions

The central premise of this thesis is that musical immersion can be understood in terms of compositional design. The listener's experience is a staged element of the compositional design. My thesis approaches immersion by focusing primarily on so-called immersive and interactive forms of media, as it is possible to show how these forms for pop music and media alter the perceived relationship between the performer and the viewer. Further, the vehicle for such change comprises a variety of new approaches to comprehending the staging of both the performer and the viewer. This argument generates several questions: primarily, how can the construction of immersive and interactive pop music multimedia inform our understanding of immersive musical experience in general? Following this, I ask: how does immersion affect perceptions of spatiality, compositional design, staging, performativity, and identity; what is the relationship between the performer and listener in popular music; and what does immersive staging imply for the subject positioning, subjectivity, and agency of the artists and listener?

While these research questions point to a strong focus on immersive media, I contend that the results of this research point to something more general about pop music and music video, which is that the experience of the listener of a pop track is integral to compositional design, and not simply an effect of it. During this thesis

I propose the term immersive staging to frame analyses that demonstrate how listener experiences are staged and compositional. While immersive media offer up a readily demonstrable case for immersive staging, I have attempted to demonstrate that the methods and frameworks on offer throughout my work are just as valuable when applied to stereophonic music recordings and 2D[3] music videos. By considering immersive staging in pop music recordings and videos, one can unpack the features of the pop score that allow fans to be immersed—not only in 3D sound, but more generally in the performances and personae of pop artists.

Aims and Objectives

This research is motivated by several aims and objectives that attempt to:

- Expand musicological approaches to analysis through studying immersive and interactive audiovisual media.
- Catalyze a musicological discourse on immersive and interactive media.
- Understand the compositional elements that contribute to immersive experiences of popular music.
- Problematize musicological discourses on staging by further considering the effects of immersion.
- Describe how the listener experience is compositionally designed through staging the listener.
- Problematize the diegesis of pop music video and how narratives are affected in VR and 360° music videos.
- Discuss how studying immersive music is relevant to the study of 'traditional' forms such as stereo music and music video.

The overarching aim is to show the various ways that the listener is a staged element of compositional design. A main aim of this thesis is to catalyze a dialogue

---

[3] By '2D' here, I am not referring to the content of a music video (and especially not to its quality!), rather, it is in reference to the flat screen that it is viewed on, such as a television, computer screen, or smartphone. This is by contrast to 3D video, where the image is larger than the field of view of the display, requiring the viewer to move using a VR headset, or through interaction such as in a 360° video. Terms like '2D' and '3D' are unendingly clumsy, in particular in discussions like this which mix technological concepts with metaphorical descriptors. This is discussed in more detail in the section titled 'Technology, Diegesis, and Aesthetics' (p. 29).

within popular musicology about newer forms of immersive multimedia and their relationship with popular music. These formats, which include surround and 3D audio, virtual reality, music in video games, 360° videos, and so on, seem to be more integrated into everyday multimedia consumption. This takes the form of 3D home theater and smart speaker technologies, binaural music in headphones (such as Atmos 3D on Tidal HiFi, Amazon Prime Music HD, and Apple Music and the recent implementation of 'Spatial Audio' features in Apple Music and Apple and Beats branded headsets[4]), pop music videos and albums in virtual reality, and the inclusion and growing popularity of 360° videos on social media like Facebook and YouTube. While I do not advocate an entirely progressive narrative about these formats, I contend that as it becomes more typical to experience and engage with pop music in a variety of immersive and interactive ways, it also becomes more important that these experiences are examined with the same critical gaze we offer to stereophonic music. Towards this effort, my hope is that the theories and methods on offer here can effectively demonstrate frameworks for analyzing and understanding pop music in these formats.

My objective is also to expand the discourses around staging and compositional design to include hermeneutics as part of my concept of musical immersion. While hermeneutics has its basis in mainstream musicology[5] I believe that interpreting the experience of becoming absorbed in performance offers insight into the ways artists may stage their personae, while simultaneously hinting at how listeners can interpret their own subjectivities in the pop performances they hear and watch. Moreover, I think that this frame provides the analyst with a novel way of considering how the listening experience is itself compositionally designed, and how the listener might be staged in pop music productions. Extending existing concepts of compositional design, I propose the idea of a pop music diegesis, a term I have coined to describe how the stories of audiovisual pop music are formed (see Article 3, p. 119).

As such, I hope to demonstrate how immersion and interactivity are relevant for audiovisual pop music in general, not only in the aforementioned 'new media' formats, but also in stereo music and music video. As I have stated in previous work, immersive formats are not required for immersive experience, and one can

---

[4] https://www.apple.com/newsroom/2021/05/apple-music-announces-spatial-audio-and-lossless-audio/
[5] For example, through the notion of ecological perception as posited by Eric Clarke (2005).

easily find immersive experience in any kind of media. Ultimately, these are hermeneutic phenomena, and while the technologies of immersive and interactive media[6] offer and easy-to-demonstrate case for how immersive experiences may be designed compositionally in pop music, I want to emphasize that this is the case in all forms of pop media. A goal of pop music, from songwriting through music production, distribution, and marketing, is, after all, to get the listener 'hooked'— to create the experience of absorption that appeals to our tastes, ideologies, and ambitions as human beings. Thus, it follows that this experience is not only an effect of popular music, but a goal of compositional design.

Structure

This thesis is article-based, comprising an introduction chapter and four articles. Throughout this chapter, I introduce the research theme and questions by accounting my own position as a researcher, as well as establishing the theoretical and methodological premises for what I describe as a 'musicology of immersion'. This entails a detailed overview of the scholarly field upon which this research builds, involving a discussion of the hermeneutic strategies that constitute the bulk of my methodology. Following the introduction chapter are summaries of the articles and chapters (p. 59), the bibliography for this chapter (p. 61), and finally the articles and chapters themselves (beginning on p. 69).

---

[6] Importantly, immersion and interactivity are not necessarily analogous. However, as I argue later in the thesis, interactivity is often a main factor in immersive experiences, and some of the formats I have studied, such as VR, are interactive media.

Popular Musicology

Methodology

The primary methods of my research stem from popular musicology and are hermeneutic and intertextual, contextualized by a discourse that takes as its starting point that musical meaning commences with subjectivity. Epistemologies that include interpretive methods are bound to receive criticism from the many within musicology who are skeptical to them.[7] Throughout, I have contextualized this thesis within popular musicology, a discipline which I believe is best defined by Derek Scott:

> Popular musicology… embraces the field of musicological study that engages with popular forms of music, especially music associated with commerce, entertainment and leisure activities. It is distinct from 'popular music studies' in that its primary concern is with criticism and analysis of the music itself, although it does not ignore social and cultural context (2009, p. 2).

Since my focus falls on what music means, I turn to a hermeneutic approach that combines formal, sonic, and audiovisual analysis with social and cultural context. While the field of popular musicology is not defined as such by close readings or textual analysis, they are a dominant for of analytical research. Accordingly, my starting point is that musical meaning can be understood in terms of musical and cultural codes and signs, and that the interpretation of these codes is highly subjective and dependent on the listener's social and cultural context (See Brackett, 2000; Hawkins, 2002; Middleton, 1990; Scott, 2009; Tagg, 1982, 1987).

Allan Moore has stressed that any study of popular music, which is a study of the music itself, must begin with interpretation, both motivationally and methodologically, since "the reason we (communally) go out of our way to experience music is simply in order to have been part of the experience of music" (Moore, 2003, p. 6). Thus, the root of musical experience is subjectivity, and that while empiricism certainly is important study of subjective experience, so too is the phenomenological textual analysis of artistic works. While the study of other

---

[7] Concerns around the merits of close readings are not new (or exclusive) to musicology or popular music studies. For example, Kramer and Tomlinson's 1993 debate in the journal Current Musicology (Kramer, 1993; Tomlinson, 1993) over the emergence of a postmodern musicology and the merits of artistic criticism (as opposed to a mainly ethnographic approach) is indicative of a divide that still exists.

musical traditions may draw upon established canons and formalized notational paradigms, popular music is first and foremost about experience, and the questions we generate in analyzing it point to this. Why do some musical experiences denote pleasure and others pain (Burns & Lafrance, 2017; Danielsen, 2006; Hawkins, 1997, 2009; Whiteley, Bennett, & Hawkins, 2004)? How do listeners understand their cultural role in their interpretations of meanings of recorded songs (DeNora, 2000; Eidsheim, 2015, 2019; Frith, 1996; Negus, 1999; Negus & Pickering, 2004; Street, 2011)? And, how does the visual and intertextual experience of music function to create an experience that seems to be more than the sum of its parts (Hansen, Askerøi, & Jarman, 2021b; Hawkins, 2002; Scott, 2009; Simon Frith, 2012; Zagorski-Thomas, 2014)?

I acknowledge numerous strong claims throughout this thesis and its articles that pertain to musical immersion and interactivity. In other fields, such as cognitive psychology, discourses on immersion and interactivity are on the cutting edge of empirical research, so the question remains: why is it relevant to study these phenomena hermeneutically? On the whole, I argue that like all other forms of musical experience, immersion and interactivity are subjective on multiple levels. This means that different listeners have different experiences of music. As a researcher, I identify as a listener, and argue for my own competence as a listener. My background both practically (in music making, production and recording) and scholastically has contributed to a heightened competence[8] as an analyst that lends weight to my analyses and interpretations. Perhaps more importantly, the interpretation tends to feel less important than the methods and frameworks by which it came about.

One compelling argument for rectifying this is the application of popular musicology. My belief is that any empirical or ethnographic study of popular music and its effects must at some point be founded on hermeneutically derived principles or hypotheses. Notably, any ethnographic study that attempts to get at what musical stimuli 'mean' to participants, for example, is definitionally

---

[8] Several have argued for 'competence' as a methodological backdrop to analysis. For example, Gino Stefani distinguished between "cultivated codes" and "popular codes" that signify musical competence, wherein the cultivated codes signify musical practice with higher class cultural capital (i.e. classical and contemporary art music) while popular codes signify the unified cultural apparatus of the masses (Stefani & Fiori, 1984). Middleton has argued that it is the role of the musicologist of popular music to "look both ways, living out the tension" (1990, p. 123) between these two distinct competencies, a task which I have attempted to embrace here.

premised on the assumption: what about the musical text and how this lends itself to meaning? An empirical approach that bypasses the listening subject by micro-analyzing the imperceptible subtleties of beat placement or tuning to infer what defines rhythm or harmony in different musical genres assumes at the outset that genre is something with a definition that lies external to the music, and assumes that these small differences are ultimately constitutive of something meaningful to people, but similarly fails to address what exactly that intermediary thing is. Anne Danielsen, who has studied microrhythm in the construction of groove in several genres of popular music (Danielsen, 2006, 2015; Danielsen & Hawkins, 2020), has employed empirical methods of technological analysis, such as waveform and sonogram analysis, insisting that the application of these tools to hermeneutic analyses is what makes close readings tangible.

While empirical and ethnographic methods are at best critical, it is my conviction that they cannot replace textualism; on the contrary, they are entrenched within this. This has been emphasized by Stephen Blum, who reminds us that "whatever we write about music is informed (in more ways than we can recognize) by our responses to works, genres, theories, performances, performers, and to many other factors, some of which we treat as 'extra-musical'" (Blum, 1993, p. 41). Similarly, in their introduction to *Popular Musicology and Identity*, Kai Arne Hansen, Eirik Askerøi, and Freya Jarman have insisted that the "integration of music analysis with an interdisciplinary mode of interpretation is imperative for unearthing connections between the musical details of composition, production, and performance, and issues of broad sociocultural significance" (2021a, p. 3).

My interest therefore lies in how experiences of popular music are designed and can employ formal and interpretive methodologies to both describe and explain experiences of musical immersion and meaning. The research I have undertaken within the realm of popular musicology is by nature interdisciplinary insofar as the philosophical and scientific framings I turn to often arise from other fields such as sociology, psychology, literature studies, media and film studies, anthropology, and so on. In considering the musical experience as central in such an approach, I position myself amongst a large corpus of work within popular music research. In recent years, several authors within popular music studies, including musicology, have made significant contributions that are based partially or entirely in hermeneutic methods. Springing to mind is Philip Auslander's book *In Concert: Performing Musical Persona* (2021), which makes the strong case for hermeneutic approaches to performance analysis in audiovisual pop music and

which is supported throughout by close readings from several popular music genres. Importantly, while many studies that analyze pop music audiovisually arise from tangential disciplines such as film and media studies, Auslander's approach is explicitly contextualized within popular musicology and implements primarily an interdisciplinary musicological framework.

Lawrence Kramer, whose theories intersect with popular musicology, reminds us that there is a subtle but important distinction to be made between interpretation in the general sense and interpretation in the hermeneutic sense, which he calls 'open interpretation':

> Open interpretation aims not to reproduce its premises but to produce something from them. It depends on prior knowledge but expects that knowledge to be transformed in being used. Open interpretation concerns itself with phenomena in their singularity, not their generality. It treats the object of interpretation more as event than as structure and always as the performance of a human subject, not as a fixed form independent of concrete human agency (Kramer, 2011, p. 2).

In *Studying Popular Music*, Richard Middleton asserted: "musical 'meaning' cannot be limited to translatable signification. In music we look not only for understanding but also enjoyment" (Middleton, 1990, p. 247). While Middleton's work pre-dates the coining of the term popular musicology and its independence as an academic discipline, it is nonetheless highly influential, arguably foundational, within it. Extending Middleton's theories into the analytic domain of audiovisuality and identity constructions, Stan Hawkins would insist that "listening is an important part of visualizing the pop music experience, where mannerism, gestures and peculiarities of the body denote pleasure, sometimes pain, with a wish to entertain" (Hawkins, 2016, p. 2). In such instances the argument for meaning and its roots in pleasure emerges, intertwined with taste and preference in pop music.

Ideologies of taste, pleasure, value, and entertainment are driving forces of meaning in popular music for listeners, and conversely, they are drivers of criticism from social and cultural elitists who would relegate popular music to a second-class status in musicology (and more broadly, pop aesthetics generally in the humanities).[9] Part of the elitism on display is the idea that the 'true meaning'

---

[9] See Brackett (2016); Frith (1996, pp. 3-8); Middleton (1990, pp. 57-60); Walser (1993, pp. 3-7).

of a musical text is the one claimed by its author or claimed to be true based on an interpretation of the author's intention. In Moore's words, this approach has served to only "divide those listeners who understood the meanings of the great works, from those who did not, or apparently could not" (Moore, 2013, p. 9). Grand narratives of aesthetic meaning miss the mark completely, since at the heart of so much music is the idea of broad relatability—popular music is a phenomenon of the masses.[10] In other words, framing the question of "what does this song mean?" as having an academically provable answer can seem nonsensical. A pop song does not possess a singular meaning; rather, pop songs are constantly and ecologically open to interpretation by listeners. Thus, any notion of a complete story is told through an interdisciplinary and intertextual approach that includes hermeneutics, since the alternative will derive only a limited understanding of popular music that excludes the people who listen and find meaning in it.

Hermeneutic analyses commence with the fundamentally ecological and personal state of listening.[11] Worth emphasizing is that all interpretations are ultimately personal and subjective to the experiences and background of the listener/analyst. Yet, the risk in interpretive methods can well be the part of the analyst, where subjective interpretations of musical details are taken as universal truths. As Hawkins warns, "(a)s a guarantor of meaning, the musical structural detail is constantly threatened by misprision and is anything but assured. What I mean is that there is always a sense of legitimacy in one's own brand of hermeneutics that seeks to validate the means of one's craft" (Hawkins, 2001). Hence, the analyst's role is not to provide the reader with a definitive answer or indeed to function as some corrective. On a similar note, Moore argues that the goal of hermeneutic analysis is not to tell the reader what a song means, but "to explain the means by which songs can mean" (Moore, 2012, p. 3). Music analysts always risk the peril of engaging in a prescriptive manner, and as popular musicologists have identified, such a hubris ultimately reinforces rather than challenges hegemonies of race, class, gender, sexuality, ability, and ethnicity, and

---

[10] For an extensive overview of what popular music is and how it is defined, see Middleton (1990, pp. 3-7). This statement could be seen as a "technologico-economic definition" for popular music, which relies on its dissemination through mass media (ibid., p. 4).

[11] See Clarke (2005), *Ways of Listening*; DeNora (2000), *Music in Everyday Life*; and Kraugerud (2021), *Come Closer*.

simultaneously undermines our capacity to describe and understand music culture.[12]

This brings me to an important caveat, something that runs throughout this thesis and its articles and chapters, which refers to the entity of "the listener" or "the viewer." In essence, the listener is an abstract and hypothetical entity who presumably shares characteristics with myself as the analyst (who's competence as a listener I have already argued for), and the general listening public. Wherever possible, I have supported my claims about 'the listener' and 'the listening experience' through not only a rigorous and interdisciplinary hermeneutic methodology, but also intertextual references, citations to music reviews, and other forms of public discourse. Thus, while I have not engaged in an empirical audience research methodology, I maintain that by speaking of the listener and the listening experience in this way allows me to abstractly distill my own interpretations with those that I might imagine are possible and those that exist within public discourse as a single entity.

In Locating the Pop Score

Throughout the thesis, I have specified that the objects of analysis are pop music recordings and music videos in their many formats, and in analyzing them have labelled them as both 'pop scores' and 'musical texts'. It is worth emphasizing that this labelling is done with the intent to build on Hawkins' concept of the recording as the pop score (2002, pp. 29-30). I concur with Hawkins that the pop score refers not only to the notational parameters, but also those features captured by stylistic and technical codes such as sound (record timbre, recording and production techniques, beat, groove, etc.), performance gesture, spatiality, audiovisuality, and so on (ibid., pp. 11-12).[13]

In the introduction to *Reading Pop*, Middleton has extensively problematized the notion of the text through a historiographical description of its ontological formation in the development of a critical musicology (2000, pp. 1-19). The terminology of the text arose out of semiological approaches to popular music study where the analyst attempts to escape the "notational centricity" (Tagg, 1987,

---

[12] See Brackett (2000, pp. 19-21); Frith (1996, pp. 3-8); Hawkins (2001); Scott (1990, 2009)

[13] The ontology of which parameters exactly constitute the score has been widely debated, and I delve into a deeper discussion of this later, in the sub-section entitled 'Audiovisuality' (p. 25)

p. 28) of traditional musicological approaches by defining analysis through reading the multitude of 'texts' that pop music generates (Middleton, 2000, p. 5).

Thinking about the pop score signals an attempt to gain back some middle ground, where the notion of the score, rather than being replaced by the notion of the text, is granted an expanded ontological basis to include those parameters (namely, stylistic and technical codes) that are left out of traditional score study (Hawkins, 2002, pp. 3-12). Thus, the pop score can be perceived as endemic of pop texts—it is the very parameters that lie behind interpretations of the multitude of texts that are generated in pop music. This subtle analytical turn back to the notion of the score is what allows for consideration of compositional design (ibid.), since this middle ground between the relativism of textualism and the structuralism of score study, in my view, grants some agency back to the music's creators. In other words, musical meaning is dialectical—it comes about through both the structural construction of notes, sounds, and codes (the pop score) and its interpretations by listeners (pop texts).

An effect of the influence of popular music scholars, and admittedly an ideological motivation on my part, is that I am interested in studying 'mainstream pop' or commercially driven texts; that is, music of the genre (or genres) which is in current or very recent public discourse, that is commercially successful and appears on various worldwide charts (such as Billboard's Hot 100) and popular playlists (such as Spotify's Today's Hits). In arguing for the serious study of commercial music in the academy Hawkins has been adamant, "If there has been one main agenda of critical musicology it has been the dismantling of the canon, its formation and the set of ideological values that have historically legitimated its study" (Hawkins, 2012, p. 3). In looking to the repertoire of popular music studied, this goal might seem unattainable as rock music has continued to dominate as the focus of much musicological study, although it has not been a mainstay of the pop charts in many years. In sum, by engaging with mainstream pop, I remain sensitive to the problems of canonizing popular music.

Technological Considerations

An underlying theme of my research is the role of technology in music production and consumption. As Paul Théberge insists, "any discussion of the role of technology in popular music should begin with a simple premise: without electronic technology, popular music in the twenty-first century is unthinkable" (2001, p. 3). Arguably, music technology is popular music's primary mediating

factor, not only in terms of music recording and production (for example the technologies of the recording studio and the digitalization of musical sound), but also in the technologies of music dissemination and consumption. These mediating factors and their implications on popular music aesthetics are of primary significance. Thus, it is worth exploring how questions of technology have been addressed within popular musicology to this point.

In the introduction to *Critical Musicological Reflections*, Hawkins accounts for a series of meetings and conferences in the early 1990s in Sheffield, UK that led to the establishment of a critical musicology forum (Hawkins, 2012, p. 5). For one of these meetings in 1993, a critical musicological charter was drafted, which aspired to numerous goals, including: "explorations of the multiplicity of music's contemporary functions and meanings, with particular emphasis on the evolution of new technologies within late twentieth-century post-capitalist cultures" (ibid., my italics). Such a call to explore music technology and its impact in the late 20th century has been heeded by many. For example, Peter Wicke has written extensively about the 20th century concept of 'sound' and how the ideology of sound quality and high fidelity has come to define recording technology in the 20th century as recording transitioned from something representational to a form in itself (Wicke, 2009, pp. 147-149).[14] Théberge has written extensively about the changes in recording, production, and performance technology at the end of the 20th century, and in particular addressed the importance of the home studio, the re-definition of what it means to be a musician, and the formation of the "singer-songwriter-producer-engineer-musician-sound designer" (Théberge, 1997, pp. 221-222). In a musicological study, Ruth Dockwray and Moore traced the development of normative spatiality in popular music recordings through the soundbox, demonstrating how the "diagonal mix" came to delineate the typical sound of recorded popular music, a topic to which we shall return later (Dockwray & Moore, 2010, p. 186).

Offering a broader view on music technology, Simon Frith insists that "the technology of music simply refers to the ways in which sounds are produced and reproduced" (Frith, 1996, p. 226). He divides music technology into three distinct eras: the "folk" stage in which "music is stored in the body… and can only be

---

[14] Alf Björnberg has done a similar historiographical approach to understanding Hi-Fi culture, looking specifically at its development in Sweden between 1950 and 1980 (Björnberg, 2009).

retrieved through [live] performance" (ibid.), the "art" stage in which "music is stored through notation… [and] can still only be retrieved in performance, but it also has now a sort of ideal or imaginary existence" (p. 227), and the "pop" stage in which "music is stored on phonogram, disc, or tape and retrieved mechanically, digitally, electronically" (ibid.). What the history of recording has demonstrated in the last 100 years is a dramatic alteration of the ontology of what counts as musical performance, namely in recorded forms. Hence, recorded music has become the primary means of experiencing musical performance.

Much scholarly discourse on music and technology in the 20th and 21st century is grounded in relatively recent history with the digitization of the music industry that began in the 1990s. For example, Robert Strachan's *Sonic Technologies* begins with the "shifts in music production practices facilitated by the personal computer" (2017, p. 20), focusing mainly on the changes to music production practices that came about due to the wide availability of digital audio workstations (DAWs). Similarly, Ragnhild Brøvig-Hanssen and Danielsen's *Digital Signatures* (2016) attempts to work out the various impacts of digital technologies such as digital reverb and delay, cut-and-paste tools, digital silence, and auto-tune Arguably, the first two decades of the 21st century has seen the most significant shift in history in terms of the proliferation of technologies in the daily lives of everyday people, and the effects of this are no less dramatic in the music production and dissemination technologies of this period. It is however important to recognize that technology has been central to popular music's story since the advent of recording in the early 20th century.

Pondering over these studies, and in particular Frith's account of music technology now over 20 years later, it seems relevant to extend the framework to add another 'era', namely the social media era, where music is not only stored and retrieved digitally, but also created and disseminated through the intertextual discursive platforms of YouTube, Facebook, Twitter, TikTok, and any number of social media enterprises that constitute the places where people hear, see, and engage with popular music and culture. For example, it seems to me that Frith's 'pop' era does not necessarily account for the rise of viral TikTok dance trends, where users upload their interpretive dances to pop hits like Ke$ha's 'Cannibal' and The Weeknd's 'Blinding Lights'. Nor does it capture the discursive nature of the social media interactions between artists and their fans, such as those Auslander described between Lady Gaga and her fans through Gagavision (2021, pp. 219-221). Social media is critical for certain immersive and interactive formats,

in particular 360º music video, which is shared primarily on the platforms YouTube and Facebook. The use of 360º cameras to record and share material from concerts, rehearsals, and recording sessions seems to be increasing on social media as artists and recordists attempt to grant their viewers more and more in-depth windows into the spaces, places, and processes behind the music they follow.

One particular research community where the technologies and aesthetics of music production have been given specific attention is in the Art of Record Production (ARP), which is not only a frequent conference, but also "an online journal (arpjournal.com), a formal association (the Association for the Study of the Art of Record Production: artofrecordproduction.com) and… a nebulous but essential academic support mechanism" (Frith & Zagorski-Thomas, 2012, p. 1). Uniquely, ARP has successfully brought together recordists, musicologists, and pedagogues who are interested in studying record production in its "technical, aesthetic, and musical" forms (ibid., p. 3). As a musician with a background in music production, I have heard on multiple occasions (in particular over drinks at conferences such as the AES) the critique of musicological study that even though the goals are noble, musicologists often get the details wrong—it's clear they know about music and culture, but it is often 'cringeworthy' when reading a study that makes a claim about reverb or compression which any recording engineer can see is plainly wrong. This complaint clearly visible in the 'interlude' sections of the ARP book that give well-known industry practitioners space to comment and critique on sections of the book, such as Bob Olhsson's suggestion that Albin Zak's chapter "concentrates too much on journalistic notions of high fidelity and ignores some of the logistical and practical changes affecting music and production at the time" (p. 92).[15] Of course, practitioners tend to focus on the very thing they practice in much detail, and as such these are the areas where they are most critical. However, the importance of academic research is to acknowledge that musical meaning is not only created in the studio, but also in listening.

---

[15] Like other groups dedicated to the study of popular music, for example IASPM, ARP has managed to achieve a balance this through the confluence of the 'insider's perspective', and by welcoming research that centers not only recording analysis but also the means of music production and the pedagogics of music recording and production. For example, pedagogues such as Paul Thompson and Phillip McIntyre have made in-roads into the recording studio's creative potential for music making and music production education (Thompson & McIntyre, 2013), and made contributions to our understanding of musical creative processes in general (McIntyre, 2012; Thompson, 2018).

Virtuality in Space and Place

At this point I want to address the concept of virtuality, since in so-called immersive music technologies, the notion that the virtual can become (or get very close to) being 'as good' as the real seems a salient point of departure in the cultural zeitgeist around these technologies. In general, the term virtual refers to that which is not (or not yet) realized—it is the stuff of the imagination, without analogue in the physical environment. Sheila Whiteley reminds us that although "all music has an element of virtuality… some artists specifically incorporate techniques that encourage listeners to understand and engage their music in a virtual space" (Whiteley, 2016, p. 2). In one sense, this virtual space for recorded music can be representational. For example, Simon Zagorski-Thomas has described how digital reverberation is used in stadium rock to create a virtual 'stadium in your bedroom' (Zagorski-Thomas, 2010). In this sense, virtual space has come to stand in for a 'real' place that "plays a significant part in the way that individuals author space" (Whiteley et al., 2004, p. 3).

As the virtual is an abstraction, it is important to contextualize it within a discussion of space and place. In popular music studies, the terms 'space and place' are often taken together, since places hold great significance in much popular music, yet such real places often stand in for a more abstract sense of space (Whitely et al., 2004, pp. 1-22). For example, much of rap music is highly contextualized, often defined, in terms of place—East Coast, West Coast, Atlanta, Chicago, etc.—and at the same time these places in rap music stand in for the Black urban space, an abstraction that encompasses not only places, but the culture, norms, fashion, sounds, feelings, etc., of those who identify with this space (Rose, 2008, pp. 62-74). Place is also an important concept in the sense that the places of music's reception are important to consider when attempting to ascertain music's meanings. Tia DeNora has written extensively about the everyday experience of music and how place, i.e., the location where one is when listening, holds key importance (DeNora, 2000). For example, DeNora has analyzed how music is used in retail settings to control the behavior of potential consumers, thus blurring the interpretive lines between music, fashion, and social control (ibid., pp. 133-138).

Further, digital technologies, and in particular the internet, has blurred the boundaries between the real and the virtual through the abstractions of the digital. Shara Rambarran has problematized the complexity of the digital-virtual, suggesting that "a way of understanding these terms is to consider that our creative thoughts and imagination (i.e., the virtual) can be either transformed or nearly

transformed into reality and actuality through digital means" (Rambarran, 2021, p. 1, emphasis in original). In other words, digital media and the technologies that support them are a substrate for making the virtual meaningfully real. In public discourse, virtual spaces such as social media groups, YouTube channels, Twitter feeds, and so on, are not mere abstractions, but firm realities that are for many as real as the actual physical places with which they associate deep personal meaning.

Thus, notions of virtuality, as conjoined to space, and place, are deeply entrenched within popular music. In particular, the ontological shift of the 'real' to possibly include the digital has enormous ramifications on any epistemology of musical spatiality. To demonstrate with a short example, consider a concert performed by Lil Nas X in 2020 on the gaming platform Roblox. While not a 'game' in and of itself, Roblox is a social platform for online gaming in which all the games are made by users of the platform with Roblox's set of easy-to-use development tools. Most games on the platform are relatively simple games set in low-poly 3D worlds, similar to the aesthetics in games like Minecraft, and can be anything from a simple single-player racing game to a fully open-world massive multiplayer online (MMO) environment. In December 2020, Lil Nas X and his team created such an MMO environment for delivering the 'Lil Nas X Concert Experience', which was in total about 10 minutes long and featured four of his most popular tracks, including Old Town Road and his 2020 Christmas season single Holiday. Using motion-capture technology, the performance was literally larger-than-life, with Lil Nas X's avatar being portrayed as gigantic in relation to the player avatars who joined in watching the experience. The concerts were very well attended, with over 33 million users watching between the two live streams, and many fans took to platforms like Twitch to live stream their reactions to the concert in the game.

This example demonstrates two important aspects of the virtual-digital spatiality. For one, as with other concerts done in MMO game-type environments, the concert itself is not a replacement for an in-person live event; rather, it is a different type of performance. For example, the avatar of the performer is able to re-shape and perform both their voice[16] and body in the virtual environment in

---

[16] Nina Sun Eidsheim's *The Race of Sound* (2019) makes useful in-roads into the staging of African American voice in music. Lil Nas X's racial identity cannot be ignored as a staged element of these performances, both sonically and visually. Eidsheim's approach to analyzing the effects of the Vocaloid processing software on recorded African American vocal sound (ibid., pp. 115-150) would thus be relevant in a deeper analysis of Lil Nas X's performances in Roblox.

ways that are not possible within the real concert setting; the boundary between the audience and the stage is routinely and frequently broken; and audience members come and go as they please, interacting with one another through various modes including virtual-physical interaction via their in-game avatars, voice chat, and text chat commentary. Second, within the digital space created in this concert are a series of meta-spatialities, including the in-game space, the voice and text chat, and the real-time social commentary of popular Twitch streamers. In fact, one could have attended the Lil Nas X concert via one of the streamers, having the concert performance completely mediated through their commentary, while simultaneously being an attendant to the concert, and interacting with other viewers by enacting virtual gestures like dancing or air guitar, sending memes in chat, and commenting through live stream. Although the virtual MMO game concert may be different ontologically from an in-person live staged concert, it is not clear that, at least to the people who attend, it is ontologically different from a 'real' live experience. In other words, the virtual space in this context seems to border on the definition of a real place in the minds of viewers.

In considering immersive media and virtuality, an obvious starting point is the technologies of extended reality (XR), which are often colloquially known as VR and which includes virtual reality (VR), augmented reality (AR), and mixed reality (MR) (Greengard, 2019, p. 4). Whether or not they have actually used a VR headset, I feel that most people know what VR is, and that the platonic ideal behind its existence is to digitally transport the user into a different, virtual environment that is made real through vision, sound, and most importantly, agency. A fundamental assumption of 'reality' is that one can at a minimum choose where to look and how to move, and the VR headset replaces the visual field of the user into the virtual environment where they can do just that. The illusion can of course be made all the more effective through adaptive sound, where the sounds one hears adapt to the movements in ways that mimic sound in the real environment.

A central aspect of music in virtual reality is the virtual performer, which refers "to performers who are available to their audiences only as mediated representations, rather than in corporeal human form" (Auslander & Inglis, 2016, p. 36). In considering the above example of Lil Nas X's Roblox performance, the performer is mediated through a digital avatar in a virtual game world, and the viewer's only apparent interaction with the performer is through this mediated form. Similarly, in Björk's VR music videos, the body of the performer is replaced with the digitized avatar, and Ken McLeod has written extensively about the

performances of holograms, such as the virtual reincarnation of Tupac Shakur (McLeod, 2016). Importantly, performers in all recorded media are virtual performers in at least a minimal sense, being mediated through the sonic and visual technologies that capture their performances and being situated temporally distanced from the viewer whose present is the performer's past. Thus, both a concert of a virtual avatar in a video game and a recording of a live-in-concert singer-songwriter are virtual performances of virtual performers.

However, VR technologies, 360º video, 3D sound, do something more, which is that they place the viewer themselves within the virtual scene in a literal sense. While one can of course feel immersed in a stereo recording or music video and feel like they are 'there', these technologies seem to actually transport you there. Thus, more than a virtual performer, these formats can result in the virtual audience, wherein at least part of the listening or viewing experience is mediated through virtual augments or replacements to the viewer's own body. For instance, in Björk's VR video 'Family', the viewer is granted a set of hands, moved through the use of remote controllers, which can interact in various ways with the audiovisual scene (see Article 2, p. 93). In another example, the 360º music video 'Stor Eiglass' by Squarepusher places the point of perspective atop a cartoonish naked human body, such that the viewer who looks down will see their virtual bare chest (see Article 3, p. 119).

Temporality

Issues of temporality are central to all music analysis, since by definition music is an art of time as much as an art of sound. In musicology, there has been a long-standing endeavor to de-temporalize music by considering the object of analysis as a static entity, either reduced to the notational score or considered as a singular object within the memory of the listener. For example, the soundbox (Moore, 2001) is a visual abstraction of recorded music's spatiality, which shows a static image of the spatial construction of a mix at one particular moment, thus freezing it in time for analysis. Arguably, score study de-temporalizes music, since it reduces the temporal features of pitch, rhythm, and dynamics to static visual and textual elements that can be analyzed. Denis Smalley has suggested that in "arriving at a holistic view… I disregard temporal evolution: I can collapse the whole experience into a present moment, and that is largely how it rests in my memory" (Smalley, 2007, pp. 37-38). Compressing the entirety of music's temporality useful in analysis (and indeed necessary if we are to describe musical

phenomena in written language). However, it ignores the bodily aspect of temporality and the pleasures of music listening as it happens. Hawkins in addressing this claims, "[to] speak of 'feeling the beat' is to accept its immediacy through time and sound" (Hawkins, 2008, p. 123), while Frith similarly insists that "every clubber knows [that] to dance is not just to experience music as time, it is also to experience time as music, as something marked off as more intense, more interesting, more pleasurable that 'real' time" (Frith, 1996, p. 156).

Another way of considering temporality is through the situatedness of a musical work in historical time or in relation to the listeners subjective historical experience. In relation to subject positioning, a track is situated temporally to a listener which can reveal their experience and reading of it. Hawkins refers to this broadly as 'temporal-specific listening' whereby the recorded track "is foreclosed by temporality; its sense of being in the here and now can indeed propel us into the then and there. Yet, it can also take us back in time" (Hawkins, 2016, p. 4).

These two ways of considering time—as the immediacy of temporal unfolding within the spatio-musical experience and as the temporal situatedness of the track within the subjectivity of listeners—have not been sufficiently recognized within music analysis. In the case of the former, too often the temporal unfolding and immediacy of musical listening is taken for granted as we attempt to identify the score; and for the latter, there is a bias towards structuralism that suggests that meaning in the pop text exists independent of the subjectivity of particular listeners.[17]

Distinguishing between formal temporal aspects and the real-world temporality is a matter of music analysis. Formal properties have been addressed by many other authors.[18] In considering the 'real' passing of time, one must compress periods of time in memory into abstract singular units when reflecting on music. To this I would suggest that a movable reference frame with regard to temporal unfolding is useful for analysis—considering minute passing moments can be as interesting and revealing as compressing the verse into a single memory unit, or indeed the track as a whole.

---

[17] Notably, Danielsen has balanced these temporal threads in her analyses of James Brown's grooves, where the deconstruction of rhythm, harmony, and vocal performance serves to explicate the pleasures of feeling the funk groove (2006, pp. 75-86).

[18] For a comprehensible overview that includes analyses of popular music, see Chapter 3 of *Song Means*, which considers the temporal aspects of meter, hypermeter, phrase structure, and syncopation (Moore, 2012, pp. 51-69).

Theodore Gracyk insists that repeated listening is part and parcel of both the joy of listening to popular music and in popular music analysis, as recordings can reveal "new facets and nuances on playing after playing" (Gracyk, 1996, p. viii). William Moylan reiterates this, suggesting that repeated listening is necessary for good music analysis, as "recorded performance… allows repeated listenings and reflection, deeper examination and more personal interpretations by the listener, and discoveries of the subtleties of the music, the lyrics and the recording" (Moylan, 2020, p. 11). Listening to a recording or viewing a music video multiple times is necessary for analytic interpretation, and this is particularly true for immersive visual media such as VR and 360º videos. In these media, the entire image is never presented to the viewer at any given time, since they need to move through the space to see their surroundings. As I have demonstrated in my analyses of 360º videos, as well as in my analysis of a VR music video, these productions demand repeated viewings, each of which represents a totally unique experience where things are heard and seen which could not have possibly been heard or seen in previous viewings. Moreover, the novelty of each viewing is part of the process where the viewer's agency is staged in compositional design, a point to which I will return later. Thus, as I have attempted in my analyses, it is critical to describe not only the structural temporal elements that make music happen, but also the temporal flow of music and how its temporality shapes the pleasures of listening in real-time.

Audiovisuality

Music is more than just sound. It is performed with gesture, expression, and dance; it is a verbal and textual discourse within pop culture and media; it is mediated through auditory, visual, and haptic media; it serves to represent places and spaces for the many people who hear and watch it. As Auslander has remarked:

> (…) contra those who would claim "music is sound, and only sound is music," that the visual and behavioral dimensions of musical performance—the dimensions through which musical persona is communicated—are essential to both the production and the reception of musical sound (2021, p. 49)

A primary goal of popular musicology has been to expand the definition of the pop score and liberate pop music analysis from what Philip Tagg called "notational centricity", that is, the "tendency to use notationally recordable parameters of musical expression as a basis for the description and analysis of pieces of popular

music" (1979, p. 28). Gracyk has insisted that "the sound of the record is part of the musical work" (1996, p. 17). In today's academic discourse, the pop score has come to stand in for nearly any parameter, musical, sonic, visual, social, or otherwise, that contributes to a song's meaning (Auslander, 2009, 2021; Burns & Hawkins, 2019a; Burns & Lafrance, 2017; Collins, 2007; Collins & Dockwray, 2015; Dibben, 2013; Hansen, 2017a; Hansen et al., 2021b; Hawkins, 1992, 2002, 2020; Vernallis, 2004).

Still, some scholars have remained skeptical to audiovisual approaches to popular music study. For instance, Moore suggests that "How the artist looks is secondary for a number of reasons… visual image is more readily accepted as constructed than aural image… whereas sound appears unmediated" (Moore, 2012, p. 101). On the contrary, as Auslander demonstrates in his analyses of Lady Gaga and Nicki Minaj that pop artists are masters of constructing seemingly unmediated visual imagery within the context of social media (2021, pp. 207-226). On the flip side, the general public is clearly aware of music's staging, particularly visible in the vitriolic public discourse around the use of technologies such as AutoTune in pop recordings. These forms of audiovisual intertextuality highlight for the viewer not only that whether the audio and visual images are 'constructed' or not is completely unclear, but also their staged candidness seems to draw attention to the high degree of technological mediation present in the sonic text through comparison.

In the study of audiovisual contexts, for example in pop music videos, television and film, live musical performances and their staging, and so on, there have been useful approaches from a variety of disciplines. Identifying primarily with popular musicology and critical musicology, I am acutely aware that there have been scholars within these fields who take seriously audiovisual pop music (discussed a bit more below). Importantly, audiovisuality is also a major part of film and media studies, and in recent years more scholarship has been given in particular to music videos from this disciplinary perspective. For example, Auslander has studied music video and live popular and rock music performances in depth. Similarly, Mathias Bonde Korsgaard has tackled music videos from multiple angles and referred to himself wittingly as a "media scholar who often

finds himself in the company of musicologists."[19] In his book Music Video After MTV, Korsgaard highlights the interdisciplinary nature of studying music videos, suggesting that a new discipline, audiovisual studies, may be relevant to their future study (Korsgaard, 2017). In the way Korsgaard has defined it, audiovisual studies fits into the field of popular music studies. However, in contrast to his approach my methodological basis is ultimately musicological. Anders Aktor Liljedahl has asserted that analyses of music videos have mostly privileged the visual (Liljedahl, 2019, pp. 168-169), something which Korsgaard admits when he says that in analyzing music videos, he "probably devotes more time to the visual than to the aural" (Korsgaard, 2017, p. 9).

Certainly, audiovisuality in pop music has become more studied in recent years, evidenced in part by several anthologies. Lori Burns and Hawkins' *Bloomsbury Handbook of Popular Music Video Analysis* (2019b) is seminal in this regard with chapters on every aspect of pop music video. Several chapters of Burns and Serge Lacasse's anthology *The Pop Palimpsest* deal with audiovisual intertextuality, including Burns and Alyssa Woods' chapter on Hip-Hop intertexts (Burns & Woods, 2018) and Hawkins' analysis of the Eurythmics 'I Need a Man' (Hawkins, 2018). Several chapters in *The Oxford Handbook of Sound and Image in Digital Media* (Vernallis, Herzog, & Richardson, 2013) address popular music videos, as did several contributors to *The Oxford Handbook of Music and Virtuality* (Whiteley & Rambarran, 2016). And many references can be found in journals in recent years, such as *Music, Sound and the Moving Image,* which was founded in 2007 and has featured several important contributions in audiovisual popular music (see, for example Burns & Watson, 2010; Jirsa & Korsgaard, 2019; Korsgaard, 2019; Liljedahl, 2019; Perrott, 2019; Vernallis & Ueno, 2013). Lastly, it is important to note the importance of the relatively recent increase in interest in video game music to the discourses on music and audiovisuality. Arguably, this has been spurred by Karen Collins' call for change in studying music and the moving image (2007) and through her book *Playing with Sound* (2013), and is evidenced by the growing number of publications with this focus and through the establishment of the *Journal of Sound and Music in Games* in 2020.

---

[19] This was said in a seminar held at the University of Agder to evaluate this thesis at the 90% progress point in June 2021. Thanks are in order for Mathias, whose comments and questions were extremely useful in the final push to complete this text.

As Carol Vernallis has pointed out, the visual is easier to describe linguistically:

> Words that describe image take precedence in all human societies over those that characterize sound… We also have fewer linguistic terms with which to describe and define a sound. We also never feel we can own or possess a sound; we cannot control and limit its boundaries, as we feel we can an image… Sound cannot often be linguistically transcribed fully (Vernallis, 2004, p. 176).

Vernallis has also insisted that music videos are a fundamentally *musical* form (Vernallis, 2008), as has Korsgaard when he suggested that music videos serve as a "musicalisation of vision" (Korsgaard, 2017, p. 85). However, Korsgaard's analyses are limited as they skirt over the very *musical* aspects that make music videos musical. I argue *qua* Vernallis that while vision is concrete and immediate, sound is etherial, immersive, all around and all the time. As it is this features of sound that music videos attempt at highlighting, in Liljedahl's words, "the musicological approach makes sense" (2019, p. 166). That said, I concur with Korsgaard's call for the disciplinary mashup of media studies and musicology as audiovisual studies, acknowledging that undertaking analyses of audiovisual texts that put sound and vision on equal footing is challenging.

Broadly speaking, my approach is geared towards analyzing pop's audiovisuality and the interpretive experiences that are contextualized within the social and cultural backdrop. I do not claim that meaning is solely structural in these texts—to the contrary, I argue that meaning is an active process created in the experience of listening wherein the performer's subjectivity is contextualized relationally to the viewers. Later, I address the role of the viewer in musical meaning from several angles, including through the ideas of pop music diegesis in the section entitled 'Technology, Diegesis, and Aesthetics' (p. 29), and through immersive staging as I consider how listener/viewer experiences are staged in audiovisual compositions.

Technology, Diegesis, and Aesthetics

Pop Music Diegesis

A central concept in my research on 360º pop music videos is that of pop music diegesis. My starting point here is that pop music videos can operate on a narrative basis and when viewed they can be read diegetically. The term 'diegesis' in film studies simply refers to the internally logical story-world of a film, and my use here is borrowed from film musicology where it is common to refer to a particular elements of a film's musical score as being diegetic, "music that (apparently) issues from a source within the narrative" (Gorbman, 1980, p. 197) or non-diegetic, meaning that the sound's source is external to the narrative. This means if the characters in the film can 'hear' a sound it is diegetic, and if the sound is 'just' for the audience, it is non-diegetic. Many have problematized this concept. For example, Ben Winters has insisted that non-diegetic music "is often just as essential to the identity of the fictional narrative space presented in film as it is in a far less 'realistic' fictional genre such as opera" (2010, p. 230). Similarly, Anahid Kassabian argues:

> Music and sound are among many aspects of a film that go into producing the sense of a diegesis. The contribute to the sense of space, of character articulation, of many things that we would label part of the diegesis. From this perspective, they are on a par with all other aspects of film, such as art direction, cinematography, and costume design. (2013, p. 91)

Following Winters and Kassabian, I would assert that the dichotomy between diegetic and non-diegetic (or indeed of 'meta-diegetic' (Gorbman, 1980)) is probably unnecessary, since music and sound are so fundamentally part-and-parcel of the construction of narrative. Indeed, as John Richardson Claudia Gorbman point out, "the very idea of a diegesis is becoming problematic, perhaps since music videos rose to prominence in the 1980s and broadened the boundaries of filmic storytelling (2013, p. 22). This is because when considering music video, the (non)-diegetic distinction creates an array of fundamental issues, namely that when the central focus of a film is the music, how does one even begin to define the boundaries of diegesis? What about when, as is often the case, the audio and visual texts tell different stories—when the music video effectively changes the interpreted meaning of a song?

Mads Walther-Hansen has coined the term 'phonographic diegesis' in an attempt to resolve the dilemma, which "emerges from the specific configuration of sounds in the recording, and is bound to the idea of recordings as perceived performances and the virtual place and time of these performances" (2015, p. 36). In his analysis his focus falls on tracks where the diegetic frame changes at a point in the song, giving the listener a window into the song's diegetic framing. In short, phonographic diegesis as an analytical concept attempts to categorize the recording's sonic mix by the perceived performance stage and by the diegetic temporality of the sounds, based on the assumption that a music recording is interpreted as a sound event happening on a virtual stage, in line with Frith's account that "to hear music is to see it performed, on stage" (Frith, 1996, p. 211). An example provided by Walther-Hansen is from the Queens of the Stone Age song "You Think I Ain't Worth A Dollar, But I Feel Like A Millionaire", which opens with the sounds of a car radio announcement of the band, followed by the thin sounds the band over the radio (complete with the sounds of the car door closing and the engine being started), before the listener is transported suddenly into the performance stage at "the moment the vocals and bass guitar enter at 1'01 where the track abruptly increases in loudness and the frequency band increases to full spectrum" (2015, p. 29). At this moment, there is not only a change in the stage of performance from the car radio to the studio, but a temporal shift from the 'present' of hearing the recorded band in the car to the 'past' as we are sucked live into the recording studio.

Although useful for sound recordings, the analytical framework of phonographic diegesis is inadequate when considering the complications introduced by music videos, which arguably restructure the narrative for the viewer. For example, although the above cited Queens of the Stone Age excerpt does not have an official music video, it would not be a necessary plot device for the video for the sonic 'transportation' to take us from the car in the present to the performance in the past. It could just as well be that the sonic change happens entirely within the head of the driver, such that the 'diegetic' shift from radio to performance is the sonic representation of becoming immersed in the song and filling in the details missing from one's poor-sounding speaker system. Or it could be that the band themselves get in the car with the driver, and hearing the introduction to their song on the radio, join in playing and singing on cue right there. These possibilities for a visual interpretation not only complicate the interpretation of this hypothetical music video, but they also show that the

distinction between the diegetic, extra-diegetic, and meta-diegetic are in fact problematic in the case of the acousmatic track, and in fact the entire sonic palette of the pop recording is diegetic, since it constitutes the entirety what might be interpreted as a story.

Returning to the music video, I wish to posit that pop music diegesis is the story being told through the confluence of the musical performance (including the lyrical story, if there is one, tone and timbre, harmonic and rhythmic styling, and so on), the visual performance (which may or may not align neatly with the musical narrative interpretation), and the interaction of the viewer who completes the story through interpretation. Sometimes, the stories told in music videos are simple and serve primarily to support the branded image of the performer. This is especially true of early MTV music videos, which consisted mostly of glossy videos of rock and pop starts performing on stage. Other times, the music video makes direct reference to a lyrical story being told, 'acting it out,' as in the introduction to Ke$ha's 2009 video 'TiK ToK' where the opening lyrical lines of the verse are performed literally. Alternatively, they can contrast or complicate an interpretation we may have formed from the sound recording, forcing us to reinterpret the story with new or additional meanings. This can be seen in The Weeknd's 2021 video 'Save Your Tears', which lyrically seems to be a lamentation about a breakup caused by the singer, but in the video, the performer appears to have a dramatic and exaggerated amount of facial plastic surgery and performs the song to a crowd of fancily dressed mannequins, which seems to serve as a critique of the entertainment industry establishment. Read intertextually, the video is possibly a statement on his anger at being left out of the Grammy nominations for his hit 2020 record After Hours.

Taking another example of how music videos complicate interpretation and diegetic framing, I want to consider the video for Maroon 5's hit 'Sugar'. Heard acousmatically, the song can be interpreted as a relatively straightforward pop love song, with an up tempo beat and light-hearted rhythmic and harmonic structure that support lyrics like "Cause I don't really care where you are, I wanna be there where you are", and "I just wanna be deep in your love" delivered in Adam Levine's swooning tenor. The music video contrasts the 'fleeting love' narrative that is typical in pop music by showing the band candidly crashing several weddings with performances of the song to the joy and adoration of the brides and grooms to be. The viewer's role in the construction of diegesis is also evident, as Levine directly addresses the viewer in the opening of the video, saying "Its

December 6, 2014. Were gonna drive across L.A. and hit every wedding we can. It's gonna be awesome… and we're late." In my reading, this serves to bring the viewer into the story, making them complicit in the surprise performances and anxious to see the reactions of the newlywed couples.

Immersive Staging

A Musicology of Immersion

In researching the VR music videos of Björk, Hawkins and I have theorized music in its immersive media format (see Article 2, p. 93). On the one hand, we argue that there is the physical aspect of immersion—as viewers we may be literally surrounded with loudspeakers or a 3D video in a VR headset. In these circumstances, the term immersion can refer to a technological frame for media that surround or envelops the user's sensory input from multiple directions. On the other hand, there is the actual experience of immersion, not a technological but a psycho-sensory and interpretive phenomenon of being lost in or completely engaged in something such that one experiences a period of intense focus.

Both these modes of immersion are indicative of a production-reception dichotomy that is often neglected in popular music research. All too frequently the production-centered analyst approaches music analysis from the perspective of the performing artist, producer, or mixing engineer, and their methods, models, and analyses are reflective of an understanding of music that uncovers the 'secrets' behind the mix. This can propel the reader toward a preference for the real rather than the perceived, and the technologies that enable experiences are prioritized to some degree over the experiences themselves. In other words, a production-centered approach is in many ways reflective of a bottom-up frame, wherein meaning in the pop score is seen as emergent from the context of its production, stylistic codes and social grounding (Hawkins 2002). A reception-centered analyst, on the other hand, approaches analysis from the perspective of the listener and viewer with approaches that seek to describe first and foremost the interpretive potentialities in a musical text.[20] This approach may not be as focused on the specific technological conditions that allow for an experience, instead considering how the musical text is received on the whole. This is a top-down frame, wherein the overall experiences of the listener takes priority in analysis and, while the descriptions may not necessarily be reflective of the processes or technologies that enabled an experience, they nonetheless describe the experience as it is perceived.

---

[20] For examples that epitomize a focus on popular music reception, see Burns (2018); Burns and Hawkins (2019a); Burns and Lafrance (2017); Eidsheim (2019); Hansen (2017b, 2019); Hawkins (2002, 2009, 2016); Kassabian (2013, 2017); Vernallis (2004, 2008).

As an example of the production-reception dichotomy, I have identified two popular graphic models for visualizing the spatial frame of stereo popular music: Moylan's perceived performance environment (Moylan, 2002, pp. 174-175) and Moore's soundbox (2001, p. 121; 2012, pp. 29-38). Exemplifying a production-centered approached, the perceived performance environment is a model for representing a stereo sound field with a view from above. The listener is positioned in the center-bottom position, and a rectangular stage is draw in front of their field of view. To the left and right of the listener are graphic representations of loudspeakers and arranged on the stage in labelled rectangles are the instruments and voices of a particular moment in a track.

Moore's soundbox similarly models the stereo sound field, but rather with a view from the front (Moore, 2012, pp. 33-34). When shown visually, the listener is not 'drawn' in the diagram—rather, the reader viewing the diagram is literally in this position. The soundbox is drawn as the front-facing view on a rectangular room, with a perspective such that the rear wall is drawn smaller with perspective lines connecting the room's corners, and sounds are represented by graphic illustrations (such as a mouth for a singer or a guitar for a guitarist). In the soundbox, the height dimension represents perceived pitch height, such that cymbals for example are drawn relatively high while the bass guitar and kick drum are drawn towards the bottom (ibid., p. 31).

Certainly, both authors are concerned with both the production and reception of music. However, their differing approaches are indicative of their methodological values. Moore's soundbox gives a first-person perspective and allows for description of the perceived 'height' of sounds due to their overall pitch relativity. The soundbox also gives less fidelity to depth given that it is a front view, and so the front-to-back depiction is given little visual space. The perceived performance environment gives most of its space to width and depth but contains no information about the frequency characteristics of sounds. Furthermore, while Moylan's model gives more fidelity to the relative width and depth of pop mixes, it does not necessarily account for the effects of frequency masking, where sounds placed in similar positions in a mix at different or competing relative volumes or distances can make them difficult or impossible to perceive at times. In short, Moylan's model demonstrates the 'reality' of a mix with a high level of detail, allowing the reader to observe the specific spatial layout of a moment in a mix. It contains perceptual characteristics to be sure, for example that the speakers are not at the maximum front or side positions demonstrating how creative panning can

create the effect that sounds are 'outside' the speakers or closer than them, but overall is useful in a more total description. The soundbox, by contrast, is almost entirely perceptual—it is shown in first-person and details the more interpretive and metaphorical aspects of one's encounter with a track. My intention is not to suggest that one of these approaches is more valuable than the other but rather to demonstrate that the choice of a bottom-up or top-down view on pop music permeates musicological inquiry and reflects the goals and values of the author. Musical meaning is first and foremost about subjectivity—how one interprets music is ultimately a result of their intentionality towards it. While understanding and describing the methods of music production, recording, and mixing are certainly helpful in guiding analysis, the role of subjectivity in the experience of music cannot be understated.

The discourses around immersion and immersive media are often caught between production- and reception-focused approaches. What is important to note is that just as the concept of space in recorded music is made of both the actual and metaphorical notions of space, so to is the concept of immersion made of both the conditions that enable immersion and the experiences of immersion. Immersive media does not guarantee immersive experience, and immersive experience is not solely derived from immersive media. Those who have experienced such media may recall times in which it completely failed to grab their attention, and anyone can easily recall a time when they have found themselves completely immersed in something such as reading a book, cooking a meal, or walking in nature. So, in activating the terms 'immersive audio' and 'immersive and interactive media', it is important to remember that these terms normally refer to media formats, not necessarily experiences.

The experience of immersion is more ecological and subjective than Moore and Moylan suggest. Yellowlees Douglas and Andrew Hargadon, for instance, have shown that immersion is often viewed in terms of the 'flow' state (2000), which has been described by Mihaly Csikszentmihalyi as the state of complete absorption in activity (1990). Flow requires both immersion and engagement, where immersion is defined as "being completely absorbed within the ebb and flow of a familiar narrative schema," and engagement is the viewer's ability to recognize a work's overturning or conjoining conflicting schemas from a perspective outside the text (Douglas & Hargadon, 2000, p. 154). Considering this definition with regard to music, it could be said that immersion is the aspect of flow in which a listener is absorbed completely into the music and its narrative, while engagement

describes the ability of the listener to contextualize the music within a broad and personally relatable intertextual framework. In other words, flow in popular music requires the listener to be easily familiar with the musical and narrative structure while relating that structure to other works and to their social, cultural, and temporal situation.

The flow state is useful and important in analyses of immersive music. For example, in virtual reality experiences Jacquelyn Ford Morie refers to the "bifurcated body", that is, the simultaneous knowledge of one's bodily existence within and without of the virtual world (2007, p. 128). Here, the concept of flow can help to explain the moments within VR experiences when one loses sense of their bifurcation and temporarily feels completely absorbed within the virtual. However, this thesis in general is scoped down to immersion rather than flow, because of the primary focus given to the musical text. Ultimately the findings I present are supported through close readings, and it is important to remember that where I experience flow may be different than another listener. While the identification of intertext can close the gap of engagement, given that intertext can constitute to a degree the range of possible familiar schema I share with another listener (Lacasse, 2000a, pp. 36-37), my main contribution is to look at the effects of immersion.

So, what is the effect of staging in immersive popular music on experiences of immersion? There are several angles from which to approach this question. First, the possibility for audiovisual elements to be placed around the viewer has a physical implication for the size and shape of the stage. Second, this reconfigured stage necessarily centers the listener and her experience such that her presence might be interpreted as being part of the composition. Following this, staging listening experience implies embodiment.[21] Given that immersion is contingent upon the deep connections between the self and the music, this potential for embodiment carries significant implications for immersion, a point to which I shall return later.

Staging and Production

Inherent in my research questions is the issue of staging—how do new media technologies reconfigure the pop stage, how do artists use technology to stage

---

[21] See Eidsheim 2015, in particular chapter 5, pp. 154-185.

themselves, and how do listeners perceive their relationship to the stage and to the artists who perform on it? In the articles that constitute this PhD dissertation, I have broached many different aspects of staging in popular music production, performance, and reception, and in this thesis one of my goals is to present a thorough argument for how immersive pop music multimedia is staged differently. Here I draw attention to the act of staging in terms of musicological inquiry. To this end, my work is engaged with staging of pop music and media in an immersive sense, where I draw together the connections of staging with immersion, identity, subjectivity, and performativity.[22]

Immersive staging is a framework for understanding how the perceived relationship between the performer and listener is mediated through technology, performativity, audiovisual composition, aesthetics and other factors (see figure 1). In brief, immersive and interactive media offer an easily visible case for how artists and listeners engage in staging themselves and how the relationships between them are compositionally designed in audiovisual music media. Although I have used the term 'mediation' here to describe the way these factors impact the performer/listener relation, an equally valid analytical framework is "framing", which comes from media criticism and is used extensively as an analytic tool by Auslander (2021, pp. 3-6):

> [Framing] is used to denote the way in which the presentation (framing) of a news story, for instance, influences the content of the story and reflects the perspective from which it is told, thus shaping the underlying reality for an audience that depends on the media for information (Auslander, 2021, p. 3).

Thus, we could say that the relationship between the performer and listener is framed by technology, performativity, etc. While this is also true, I am reluctant here to use framing, as it might imply that the factors that impact this relationship have agency, which is not necessarily the case.

In order to explicate this immersive staging framework, I now enter into a discussion about staging in general: how staging is understood and developed within popular musicology, and, in turn, how it implemented within this research. Following on, I explore the staging of artists, which deals with the ways artists use

---

[22] I align my work to scholars whose work I build on, such as Burns (2016, 2018), Marc LaFrance (2013), Auslander (2008, 2009, 2021), Hawkins (1997, 2002, 2009, 2016, 2017), Whiteley (2000, 2016, 1997), Susan McClary (1991, 1993), Hansen (2017a); Hansen, Askerøi, and Jarman (2021b), and others.

audiovisual technologies to shape their performances Thereafter, I delve into the compositional design of listening experience, asking how is it that listeners are staged within music media. Finally, as the relationship between the performer and listener is implicit in staging, I consider how this relationship is mediated through technology and temporality.
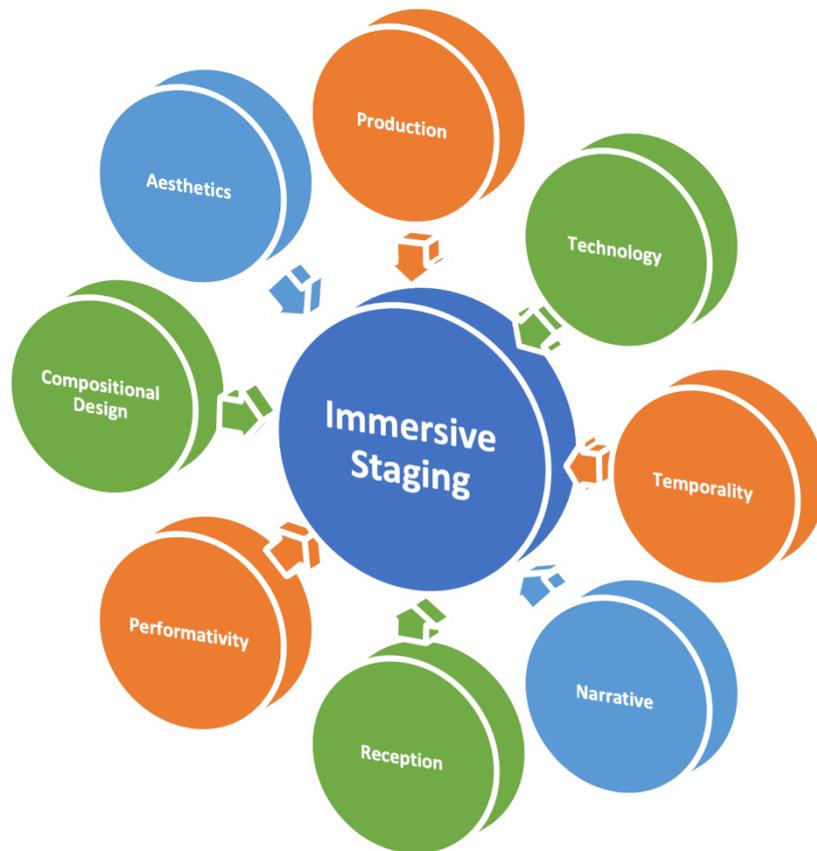

Figure 1: Immersive Staging

A number of studies within musicology focus primarily on popular music staging (Auslander, 2021; Camilleri, 2010; Dockwray & Moore, 2010; Hawkins, 2016; Lacasse, 2000b; Moore & Dockwray, 2008; Moore, Schmidt, & Dockwray, 2009; Moylan, 2002, 2020; Sandve, 2014; Zagorski-Thomas, 2010), which in general has two distinct, albeit sometimes unstated, meanings. First, the metaphor of the stage can be useful to describe how sound objects within a mix are spatially structured, which I call the physical stage metaphor. Second is the act (verb) of staging, which describes the ways that people present or are presented in performative contexts, which I call the performative staging metaphor. Importantly, performative staging is tied to notions of identity, subject positioning, and persona, since it deals with the ways personae, characters, and even listeners

38

have their subjectivities negotiated in the space of the recording. This distinction between the physical and performative staging metaphors is often subtle and simultaneous, but it is still critical to understand since in pop music analysis one runs the risk of simply describing the contents of the recording (physical) without making any inroads into the actual ways staging can communicate meaning in pop recordings.

The physical stage is a type of imaginary, empty space that is usually delineated on either side by stereo speakers and upon which the mix is built up—the spatial frame that describes the apparent location, size, and parameters of the sonic objects in a recording as well as the perceived spatiality (or spatiality's) represented. For example, Moylan's research on recording analysis and the aesthetics of popular music recordings makes heavy use of the physical stage metaphor to describe visually and textually the construction of a pop mix. In his 'perceived performance environment' diagrams, a rectangular stage is drawn with speakers on the left and right, and the discreet instruments and sounds are drawn in their relative positions on the stage (Moylan, 2002, 2012, 2020). Similarly, the soundbox (Dockwray & Moore, 2010; Moore, 2001) is a useful heuristic for visualizing the staging of recorded elements in a pop mix. An important analog to the stage metaphor is Hawkins' conceptualization of the "platform" (2016, p. 14), which "is intended to suggest the mechanism for staging production and, moreover, for archiving performance (in the form of collective social memory)" (ibid., p. 30).

Lacasse's study of vocal staging in rock music is also relevant here, since it lays out thoroughly the techniques and technologies that have contributed to contemporary rock and vocal stylings (2000b). Zagorski-Thomas' notion of 'functional staging' importantly describes how mixing decisions about certain aspects of physical staging (such as reverberation vs. dryness of particular elements in dance music, or the spatial characteristics of stadium rock recordings) are driven by consideration of the recording's intended playback environment (2010).

The performative staging metaphor is subtly different in that it is an active process in which artists and listeners express their agency through media (Burns & Lafrance, 2017; Burns, Lafrance, & Hawley, 2008; DeFrantz, 2004; Fathallah, 2021; Hawkins, 2004, 2018; Miles, 2020). In moving through the physical to the performative staging metaphors, the discourse shifts from questions such as 'where is the lead vocal panned in the mix?' to questions like 'how has the singer staged her gendered identity through performance?' Both of these questions are about

staging (and indeed about how technology mediates the composition, production, and listening to pop tracks). However, the latter is a more specific question that has the potential to address the subjective aspects of pop music production and reception.

One productive way of considering performative staging framing is imported by Auslander into popular musicology (2021, pp. 3-6). The concept of the frame is similar to that of the stage or the platform, in the sense that frames are understood through "structures of expectation" (Tannen, 1993, quoted in Auslander, 2021, p. 5), where "everything we know, from the identity of the artist, to the genre of music, to the venue where the event is to occur and beyond structures the expectations we have of the [performance] event" (ibid.). For something to be considered music in Auslander's explanation, it needs to be framed as such. Framing thus has a built-in socio-cultural element that staging does not necessarily employ, which could serve not only to describe the pop score, to to provide an ontological basis for it.

Scholars have noted that the physical and performative metaphors of pop staging are intrinsically linked. For example, the soundbox model is frequently used not only as a physical framework for describing the spatial configuration of pop mixes, but also in combination with models such as sonic proxemics to generate analyses about artists' staged personae and characters (Collins & Dockwray, 2015; Moore, 2012, p. 185). Functional staging argues that physical sonic aspects, like the reverb on clapping sounds and the dryness of drum sounds in dance music, have socially determined functions that can denote, for example, collective action in the case of clapping or shouting (Zagorski-Thomas, 2010).

By activating the term 'staging', I have attempted to describe how staging functions in many nuanced ways within popular music studies. The double meaning of staging as both a description of spatial configuration in the pop score and the metaphorical act of (re)presentation of subjectivities 'on stage' is in my opinion intentional. In understanding pop music in its immersive format, it is important to visualize how the physical stage is reconfigured in order to perceive how this reconfiguration enables changes to musical interpretations. For example, much has been written about both the norms of panning in pop vocal recordings (Dibben, 2012; Lacasse, 2000b; Moylan, 2002), and about how artists use these norms to perform their personae (Auslander, 2009; Hansen, 2017a; Hawkins, 2020). But what does it mean to the viewer when the singer performs with backup singers who are positioned behind or above them? Or when the reverberation and

delay in 3D space literally change the perceived sonic characteristics of the listening space? And how does being on the stage differ from observing it?

Staging and Immersive Media

As I have already intimated, the concept of the stage in recorded music is a metaphor, and within this metaphor the listener imagines a virtual performance space. One way of considering this virtual performance space is through Moore's soundbox, which consists of four sonic and spatial dimensions. The first is time, followed by "laterality of the stereo image, perceived proximity of aspects of the image to (and by) a listener, and the perceived frequency characteristics of sound-sources" (Moore, 2012, p. 31). As I have already discussed, Moylan has similarly proposed a spatial frame for understanding the stage, which he calls the perceived performance environment (2012). Moylan's top view diagrams depict both width and depth, but not height. Lelio Camilleri has described the 'sonic space' as the three dimensions of "localised space, spectral space and morphological space", where localized space describes the width and depth of sounds, spectral space their frequency, pitch, and timbral qualities, and morphological space the ways in which sounds operate through time (Camilleri, 2010, p. 202).

In the following section, I choose to interact with the soundbox in its various dimensions, with the aim to expand its applicability to music in extra-stereo formats. I argue that when considering surround and 3D audio in particular it might be more relevant to transform the soundbox into a soundsphere, where the listener is centered on the focal point of a sphere rather than viewing a box from the outside. The soundsphere, in short, more accurately describes the possibilities for staging in immersive and interactive media and it will be further described shortly.

The dimension of the soundbox that Moore calls 'laterality of the stereo image' is equivalent to what many call 'stereo width' (D. Gibson, 1997; Moylan, 2002; Senior, 2012), that is that in stereo, sound sources are panned along a left-to-right axis. In general, the furthest sound perceivable along this axis in any given recording represents the outer limits of the metaphorical stage—the edge of the soundbox is the edge of the stage. Moore is keen to emphasize the differences between the perception of width and distance in speakers compared to headphones, claiming that in the latter situation, "there is no distance between the sound stage and the listener – the listener is the sound stage" (Moore, 2012, p. 36).

In immersive music, the differences between headphone 3D and loudspeaker 3D are not as pronounced as in stereo. While there are of course differences in the

experience, and in some cases issues with binaural audio-only headphone 3D sound in particular, in general, headphone and loudspeaker 3D formats both center the listener on the sound stage. Thus, it is reasonable to assume that the interpretive experience will be similar between the two modes of listening. Given the choice between loudspeaker and headphone 3D, my experience is that loudspeakers offer a greater degree of fidelity, since the speakers are spaced apart allowing for the maximal spatial effect. However, in many cases within immersive multimedia, the option between headphones and speakers is non-existent or trivial for most listeners. For example, in virtual reality, all listeners will experience the immersive sound via headphones, often spatialized with the head-mounted display and its head-tracking movements. Except in research situations, it is extremely rare for VR experiences to include a head-mounted display with loudspeaker audio. Considering formats such as Dolby Atmos Music, for example, which is implemented on Tidal and Amazon Prime Music, most listeners likely do not have Atmos compatible loudspeaker sound systems, and the main mode of listening to this format at present may in fact be binaural headphone playback.

The soundbox dimension of laterality is easily transferrable into the soundsphere (see figure 2) as the dimension of directionality. While laterality considers sound objects existing on straight-line axis horizontally along the soundbox (or through the listener, in the case of headphones), directionality considers the listener in the center of the soundsphere, with sound objects able to be panned in any direction along a spherical plane that surrounds the listener on all sides. In ambisonics, a 3D format for sound recording and production, it is common to refer to the directional coordinates of sound objects using the angles of azimuth and elevation in combination with a numerical distance. Azimuth is simply the angle of the object in a circular plane around the listener, where an object directly in front of the listener is at 0º. Because of the symmetrical nature of human listening, typically the angles to the left of the listener are expressed in negative degrees and to the right in positive degrees, such that a sound exactly at the left has an azimuth value of -90º, to the right +90º, and to the rear 180º. The second angle that constitutes direction is the elevation, which is expressed in an angle rather than a unit distance in order to preserve the spherical model. Here, an angle of 0º represents a sound source that is level at the height of the listener's head, +90º is directly overhead, and -90º is directly below. So, as an example, a sound object which is slightly elevated and panned diagonally slightly to the left might have an azimuth of -30º and an elevation of +45º.
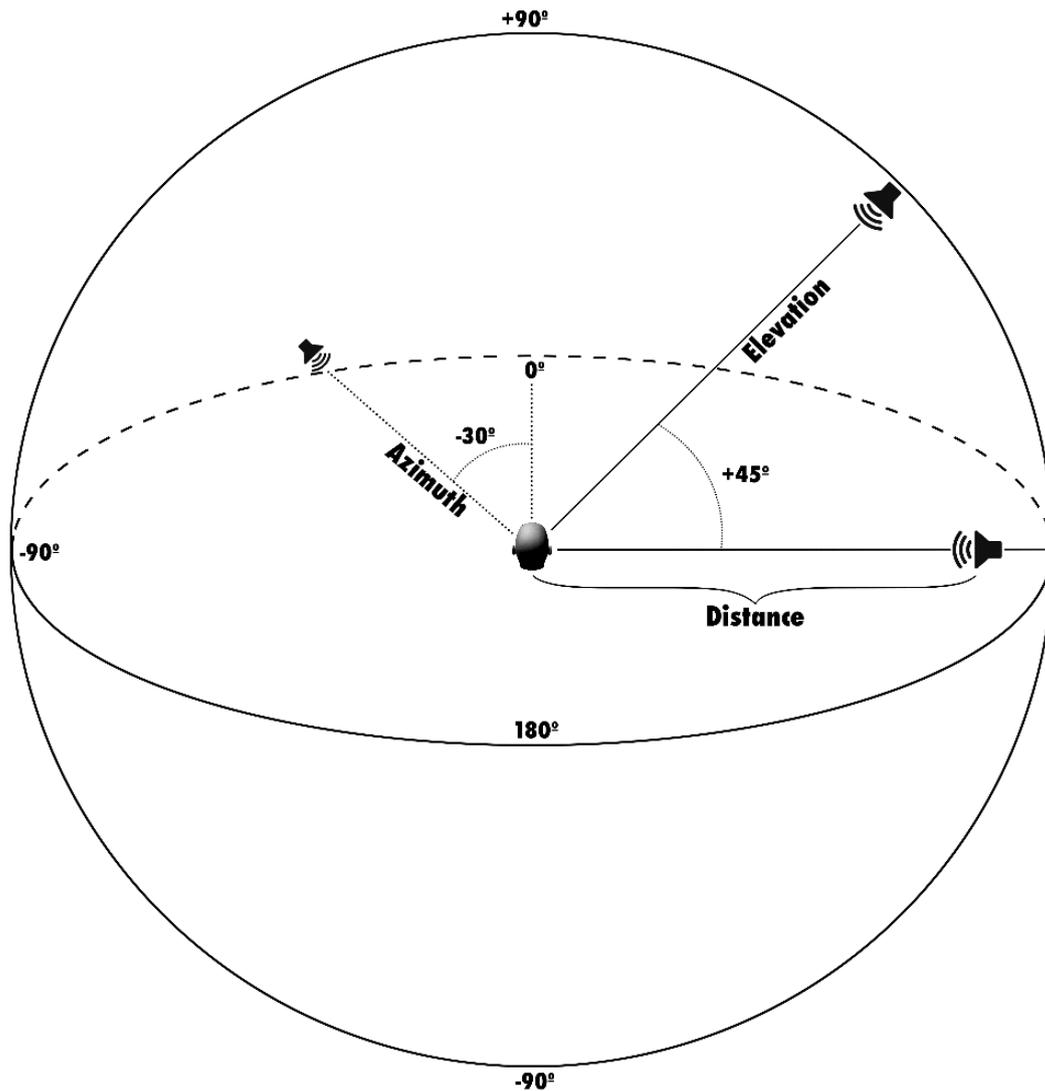
Figure 2: The Soundsphere and its physical dimensions: Azimuth°, Elevation°, and Distance

As I discussed earlier, there is some debate about whether or not the perception of height via the notational concept of pitch should be included in visualizations of spatial models. Moylan, for example, explicitly chooses to exclude it from his perceived performance environment model (Moylan, 2012, pp. 166-167), while Dockwray and Moore, Camilleri, and David Gibson choose to include it (Camilleri, 2010; Dockwray & Moore, 2010; D. Gibson, 1997). Psychoacoustics research has shown repeatedly, using various methods across decades of study, that perception of higher pitch as being physically higher is not simply a metaphor, but an artifact of the way the auditory system has evolved to perceive sound and pitch (Hebrank & Wright, 1974; Roffler & Butler, 1968; Wallis & Lee, 2015). Additionally, by testing the pitch-height effect using a broadband source (that is, a

pink-noise signal with varying equalization boosts at particular frequencies), Wallis and Lee show that this effect is more complicated than the notion that high sounds seem higher and low sounds lower, but that sounds with complex overtone structure can be perceived as higher even when their fundamental pitch sounds are lower (Wallis & Lee, 2015). For example, guitars with large amounts of harmonic distortion may be perceived as elevated via the pitch-height effect even while playing low notes. All this is even more complicated by the fact that in immersive and interactive media, sound is often literally positioned in the height dimension.

While I concur to some degree with Moore that the perception of pitch and height has large ramifications for how we experience pop mixes, it needs to be understood as more complex phenomenon. When analyzing music in 3D formats, I would argue that it is of little consequence whether the perceived height of a sound source is due to its physical panning or to the pitch-height effect (or both). Since I advocate primarily a reception-focused approach to hermeneutic analysis, what is important is whether or not sounds are perceived to have height. With this in mind, interpretations of immersion that deal with height need not be in so-called immersive formats—it is just as well for surround sound or stereo recordings to immerse us with the illusion of height created by pitch effects. Importantly, this also draws attention to the fact that mixing engineers who are aware of at least the basic principles of psychoacoustic phenomena, such as pitch-height, can and do use this knowledge to their creative advantage in creating mixes that have great impacts in all spatial directionality in any kind of audio format.

Another dimension of the soundbox relates to 'perceived proximity', which has two components: the perceived distance that a sound object is from the listener and the metaphorical proxemic function that is interpreted. Of course, these two notions are intrinsically linked, however the perception of an object's distance in the mix and what that distance means in terms of social function are two different things. At any rate, as Collins and Dockwray have insisted, proximity in recorded audio is a complex phenomenon that is constructed by many factors including microphone choice, microphone distance and angle, reverb, delay, compression, amplitude, and mixing (Collins & Dockwray, 2015, p. 54). In general, it can be said the distance and proximity in the soundsphere operate according to similar principles as in the soundbox, the only technological difference being the capability of systems to position sounds around the listener (which has already been discussed). However, surround and 3D audio complicate proximity through

44

the common use of acoustic modeling, wherein sounds are often given spatialization across the entire mix, and spatial characteristics of modeled spaces and places can be realized in all directions. For example, a sound in the front left, accurately modeled to a particular room acoustic footprint, will have its first reverb reflection in the rear right followed by reverberations across the space. As I demonstrated in my article analyzing The Weeknd in Dolby Atmos, this can create the possibility where a sound can have an incredibly dry and forward mixed sound (an intimate proxemic) while also retaining high degrees of reverb and delay through the use of the rear sonic space (see Article 1, p. 69). In a stereo mix, this effect might be achieved through the use of a side-chain compression or gate on the reverb, wherein the vocal line is left 'dry' until the end of a phrase when the reverb opens up and creates space. However, in immersive mixes, the simultaneous spatiality of the intimate voice with the large room reflection is possible without the use of these methods.

Artist Staging

Already established is that the physical dimensions of the stage are reconfigured in immersive media. It then follows that the artists, producers, and mixers who use these technologies for their work make use of these new-found spatialities to modify the performance of their identities and personae. This is critical to examine since, in pop music, the spectacle of the performer is a major part of what constitutes the interpreted pop text. In other words, as much as one can be absorbed in the ebb and flow of instrumental and rhythmic sounds, so too are we immersed in subjective identification with the self-presentation of pop performers.

Extending and systematizing Frith's account of the layered performance persona (Frith, 1996, p. 187), Auslander insists that the performer is in constant negotiation with three layers of performance, which he identifies as "the real person (the performer as human being), the performance persona (which corresponds to Frith's star personality or image), and the character (Frith's song personality)" (Auslander, 2021, p. 27). These layers, while sometimes in logical contradiction to one another, are nonetheless simultaneously enacted. In considering the immersive experience of performance, persona, and character are critical and interlinked concepts, I ally myself with Auslander in insisting that the "real person is the dimension of performance to which the audience has the least direct access" (Auslander, 2021, p. 28). I concur with Hansen that "the persona is always open to contestation and change, but still retains a great deal of continuity

over time, or in different places and situations (2017a, p. 29), and that persona and character are always co-present in pop texts. Sometimes, they are clearly distinct while at others they are essentially synonymous.

As an example, let us consider the performance of The Weeknd in the track 'Can't Feel My Face' from the album Beauty Behind the Madness (2015). This track is often read as being about the singer's (Abel Tesfaye) propensity for drug use as he personifies addiction as a woman with whom he has a complicated and dependent relationship. Tesfaye has contributed to this interpretation by being open about his casual drug use, telling Rolling Stone in a 2015 interview, "I never needed detox or anything. But I was addicted in the sense of 'Fuck, I don't want to spend this day without getting high'." While it is clear that Abel Tesfaye (the real person) is not the same as The Weeknd (the persona), the character that The Weeknd plays in 'Can't Feel My Face' is arguably that of Abel Tesfaye, the performance a complicated double-enactment which serves the function of substantiating The Weeknd as an authentic portrayal of Tesfaye as a real person. In most of his music, it is easy to read The Weeknd's performances as an enactment of the self he wants us to believe he really is, regardless of how close Tesfaye the character is to Tesfaye the real person (something we can probably never know).[23] Since by using the persona The Weeknd to enact the character of Abel Tesfaye, the artist performatively engages in 'keeping it real' as a means of self-authentication (Rose, 2008, p. 134).

While a character may be part of a performance within a particular song, the persona is an overarching image of a performer that is constructed not only within each individual piece of media, but also in the sum of all their music, videos, public appearances, media interviews, and so on. Here I turn to intertextuality, which is key to my understanding of musical personae in general. Burns and Woods demonstrated that "artists build their interests to claim power and authority within the genre, to address challenges of fame and celebrity status, and to negotiate representations of gender, race, and class within the industry" (Burns & Woods, 2018, p. 215). While it may be tempting to limit the scope of popular music personae to the ways they are constructed in compositional and sonic properties (for many, the only text analyzed), an intertextual perspective insists that the visual

---

[23] This interpretation could be viewed in terms of Moore's concept of authenticity as authentication (2002, p. 210), which focuses not on the definition per se of 'authentic' in terms of popular music, but the means by which artists attempt to achieve authenticity in an active process.

is as primary as the sonic in the construction of pop personae and the way they are staged for the viewer. Frith reminds that "to hear music is to see it performed, on stage, with all the trappings" (Frith, 1996, p. 211). In other words, even listening to an acousmatic recording is a kind of virtually visual experience, one which is made more salient when the listener carries intertextual references in their memory.

Much has been theorized about the way people understand the meanings of sounds through embodiment and visual metaphor. For example, theories of ecological perception (Clarke, 2005; J. J. Gibson, 1977, 2015) present psychological evidence that acoustic sounds are understood visually in memory—to hear a sound and understand its meaning is to visualize its source. For example, in hearing a performed guitar, one understands its meaning through imagining oneself playing one or to recall previous experiences of seeing the performed guitar. Similar claims are made in composition studies, for example by Smalley who refers to the experience of 'source bonding' in acousmatic music as "the natural tendency to relate sounds to supposed sources and causes, and to relate sounds to each other because they appear to have shared or associated origins" (1997, p. 110). Nina Sun Eidsheim has argued for a consideration of listening experience as "vibrational practice" (2015, p. 3), rejecting "the position that sound is a fixed entity and the idea that perceiving sounds depends on what we traditionally refer to as the aural mode" (ibid., p. 8). These interdisciplinary perspectives remind that meaning is fundamentally individual; the meaning one gleans from a musical text is entirely dependent on the way their experiences shape their interpretations, and readings that differ from the 'intent' of the composer are not only valid, but part-and-parcel of musical experience.

Importantly, while in this section I primarily approach embodiment in the sense of the listening experience and the formation of musical meaning, it has also been theorized in the sense that bodies are represented in musical texts. For example, Hawkins has used the concept of hyperembodiment to show how "the body's technological constructedness constitutes a prime part of the show" (2013, p. 468). In another study, Burns et al., have taken a systematic approach to understanding embodied subjectivities in lyrical and musical expression in an attempt to bridge the notions of embodiment in both the production and reception of popular music (Burns et al., 2008).

So how do artists' staged personae function with regard to musical immersion? Mainly, I argue that the audiovisual pop performance signifies a multitude of

opportunities for identification with the viewer, wherein said viewer interprets codes of identification that relate their own experiences intertextually with the persona on display. Codes of identification are aesthetic features of compositional design—either sonic, visual, or both—which the viewer recognizes as signals of a performed identity and personae that are relatable to their own identity (Auslander, 2009; Burns & Lafrance, 2017; Hansen, 2019; Hawkins, 2020). In other words, these are codes where the viewer recognizes and possibly identifies with performatives, those acts which Judith Butler (1993) insists constitute the social construction of identity. Importantly, the recognition of codes of identification need not be affirmative; it is just as well to identify antagonistically with particular codes or entire sets of codes with regard to one's own subjective experience. To immerse oneself in a performed persona is to become captivated by the spectacle of the performance and performativity of another.

One way of considering codes of signification is through the notion of proxemics, which can help us consider the perceived distance between the viewer and performer (Moore, 2012, p. 187). This is helpful since, while distance can be considered as something inherent in the quality of sound (for example, as a result of reverb and delay), it is also something interpreted by the listener. Collins and Dockwray have described in detail the various technological and methodological means by which artists can go to achieve the sonic qualities of proxemic distance (2015, p. 54), while Hawkins has shown how sonic proxemics is contingent on "an awareness of the artist's 'persona'" (2020, p. 244). As such, perceptions of proximity depend not only on physical sonic characteristics, but also the listener's interpretation of sonic, musical, social, and cultural codes.

To elaborate, I will draw on my own experience of immersion in the performance of persona. In 2002, as a thirteen-year-old, I had few friends and was quickly becoming the combination angsty musician and class clown that would come to define my personality in adolescent youth. It should come as no surprise then that one of my favorite songs from that year was Coldplay's 'The Scientist', a deceptively simple-sounding pop-rock hit in which the singer and front-man Chris Martin delivers a cool-sounding sung apology to a broken relationship. Martin, like most of the singers I idolized in my early teen years such as Radiohead's Thom Yorke and Modest Mouse's Isaac Brock, frequently sings around the break between full and falsetto voice, offering what to my ear even now is a performance of vulnerability. As a boy growing up in the conservative Midwest USA, such displays of vulnerable masculinity were socially framed as

weak, feminine, and clearly not to be emulated. Yet, regardless of the fact that I consistently attempted a (failed) performance of the macho persona, the inward, reflective swooning voices of these performers were to me utterly immersive. The sonic score offered up other codes for identification—as a budding musician its rhythmic and harmonic simplicity (a straight 4/4 drum beat that rarely changes and an ever-repeating sequence of three chords—Bm7, G, D—ending in an angsty Dsus) was easy to learn, and I spent many hours after school in the band room attempting my own private performances of the track.

At that age, I was already completely taken with technology, and the music video was utterly captivating. The narrative of the video offers an alternative interpretation of the lyrics. Entirely shot in reverse, the video features mid- and close-up shots of Martin walking backwards, the footage having been reversed, while appearing to sing in the correct temporal direction (a feat which required him to learn the song backwards and apparently took a month to learn). Close-up shots of Martin singing while walking are a mainstay of early Coldplay videos. The video for 'Yellow' for example is entirely done in one take and features only the singer walking alone on a cold beach with wet, disheveled hair, a black loose-fitting rain jacket, and baggy jeans while mouthing the words to the song. This format reads to me as a performance of pensiveness—the singer authenticates his lyrical and musical vocal expression by staging a performance that is candid and directly addressed to the viewer. Towards the end of the video, Martin approaches what appears to be a body on the ground, before entering a car with its front window shattered. Soon it becomes clear that he was the driver in a terrible car accident that resulted in the death of his supposed partner, the video meant as an apology and a wish to 'turn back time'. The partner—a woman with short, died black hair, a leather jacket which she removes to reveal a pink, midriff-exposing blouse, and a joyful personality as she is seen in the car with Martin joking and laughing—is the icon of the 'manic pixie dream girl', an irresistibly objectifiable image for a young teen nerd in desperate need of someone, preferably someone who looks like that, to understand him. Needless to say, by this moment in the video a 13-year-old version of me is silently wiping away a hidden tear.

This example attempts to illustrate how the stylistic and technical coding of various aspects of the pop score exposes subjective identification between the listener and the performance. While another viewer may identify with the recording in a different way (or indeed not at all), my claim is that the potential for identification nonetheless constitutes audiovisual codes of immersion. Ultimately,

these musical codes of identification and immersion bridge the stylistic and technical codes of compositional design (Hawkins, 2002, p. 10) through the subjectivity of the listener. In other words, immersion in performance is enabled through compositionally designed moments where the listener is afforded the opportunity to identify directly with the performer through sound and/or image. In the next section, I address this from the perspective of said listener, however, here it is critical to reiterate that what I am talking about is something that is in the audiovisual pop text itself—a musical code that is interpreted by the listener through their ecological position and carries the potential to enable experiences of immersion.

Listener Staging

Throughout this study, I make the claim that immersive and interactive pop music media stages the listener in dynamic ways. Rather than being considered as a passive observer, the listener should be thought of as a staged participant and dynamic object of compositional design. We know that listening to music is highly ecological and that listeners bring their backgrounds, experiences, temporal and physical location, cultural situation, moods and emotions, and a host of unpredictable traits to their experiences with music and music culture (Clarke, 2005). However, the listener's interpretation is also shaped by the intention of the artist, and we can say that there are intended points of view that are framed by the composition. Here I turn to the concept of subject position, defined by Sheila Johnston as "the way in which a film solicits, demands even, a certain closely circumscribed response from the reader by means of its own formal operations" (Johnston, 1999, p. 333). Applied to popular music, this suggests that there are interpretive frames which are composed and thus 'built in' to the musical experience. Considering music listening as ecological while also having subject positions may at first seem contradictory, but it is rather the opposite: it is a degree of shared background, culture, and time that enables artists in some ways to construct musical experiences that predict or demand particular interpretations from listeners and viewers.

I want to draw a distinction between subject position as I have just described and subject positioning, a verb, which, rather than referring to the particular structural frame constructed for a listener's interpretation, has to do with the ways listeners negotiate their relationship to and understanding of musical experiences. I propose that while subject positions are compositionally designed and structural

entities, subject positioning is an active process done by both artist and listener in which the music serves as a medium for asserting one's agency and identity. By staging the listener in the center in a highly engaging and interactive format, immersive popular music expands the possibilities for subject positioning to occur. When listening to 3D music, watching a VR music video, interacting with a concert through a MMO video game, or watching an interactive 360º music video on a mobile phone, the viewer is invited to express herself through her interactions with and movements through the performance stage.

So, if immersive media can more easily enable immersive experience through their spatial reconfigurations, what is it about the experience itself that allows for this? Already discussed is the way that music is designed with codes of identification, where listeners relate their subjectivity to those on display in the audiovisual pop text. However, music is more than the perceived interpersonal relationship between artist and performer. Certainly, the instrumental arrangement, spatial configuration, acoustic profile, overall balance, timbral qualities, and any other sonic and spatial quality of a pop track contributes to one's sense of immersion.

One important way that immersive media enables immersive experience is through embodiment—by staging the viewer as an active participant in the pop score, their virtual body is transported into the experience. Much has been written about the body in relation to pop music, in particular regarding the representational power of the singing voice to stand in for the body. For example, Kay Dickinson has written about how the vocoder functions "around the representational practices of the voice, of computer-made music, of femininity and of homosexuality" (2004, p. 163). Hawkins has insisted that "the staging of the voice is all about corporeal presence and active participation" (2016, p. 2). Eidsheim's study of listening and singing as "intermaterial vibrational practices" (2015, p. 3) carefully considers the body's sensory apparatus as central to musical practice, thus approaching questions of the body both "in and as performance, and as it manifests itself to us as a result of cultural construction and habituation" (ibid., p. 11).[24]

Vocal embodiment is ultimately about the body of the performer. I am keen to emphasize how the viewer's body is represented in, for example, VR. The notion of musical movement and dance is aligned to audiovisuality in pop and is encoded

---

[24] See also *The Oxford Handbook of Voice Studies* (Eidsheim & Meizel, 2019)

by embodied experience. This accounts for the social and cultural ways that music makes us move. Hawkins' theories of dance music within club settings deal with corporeal response:

> To submit to the beat is to become part of an egalitarian community entrenched in a type of religious mysticism. Stylized trends of address in club culture relate directly to the ways in which body movements interpret music in specific social spaces without any recourse to clarification through words. So, while dancers are able to focus on their own individuality, their physical motions function to establish a 'communal ethos' which, in turn, define the event, genre and context. (Hawkins, 2003, p. 100)

Similarly, Thomas DeFrantz has addressed bodily signification in dance, within the African diaspora, insisting that dance especially for members of the Black community is not only about moving and reacting to the beat, but also to communicate "performative gestures that cite contexts beyond the dance" (DeFrantz, 2004, p. 67). This is in contrast to Frith's somewhat problematic account of dance, which suggests that dance is "unnecessary movement, and end in itself rather than a means to another end… chosen for aesthetic rather than functional reasons" (Frith, 1996, p. 221). On the contrary, I would concur with DeFrantz that dance is signified in the musical score as interpreted by the listener who is compelled to move. In this way, music (and especially dance-oriented popular music) signifies the listener's body through musical structures that encourage dance, and in moving, dancing, and visualizing dance, the listener is in dialogue with their own bodily experience through these audiovisual signifieds. Indeed, the dielectic extends from the dancer on to the crowd, where "the pleasures of dancing… are about merging into a 'whole' where the emphasis falls on unity and inclusion" (Hawkins, 2008, p. 133)

From this one might conclude that dance is part of a musical function that extracts and draws in the corporeality of the listener. Hawkins insists that dance "is reinforced and enhanced corporeally" and as a result "entails an immediacy and intensity that cannot be achieved in any other manner" (ibid., p. 121). In my consideration of virtual reality, however, the body is centered in experience in this way without the 'need' for compositionally designed codes for dancing. The reconfiguration of the stage around the listener itself is the centering tool in this kind of musical experience. In this way, subject position in 3D sound and virtual reality is not only an interpretive concept, but a spatial one. In some contexts, this

opens for bodily representations of listeners which are not as open for interpretation. When considering Björk's VR video 'Family' (see Article 2, p. 93), Hawkins and I observed that the viewer is granted a set of virtual hands which are manipulated via Oculus Rift controllers, and which can be used to interact with the audiovisual scene. Stylized digitally, the viewers' hands operate in the same way as Björk's own semi-translucent and psychedelic pastel colored corporeality; when pressing the controller's trigger button, the viewers' hands move in the same swirling 'conducting' gestures that Björk herself employs throughout the performance. In this way, the viewer is granted not only the role of observer, but also that of a participant, mimicking the movements and gestures cued by the main performer.

Effectively, this sums up immersive staging; it demonstrates how immersive media reconfigure the stage for popular music in ways that dramatically impact the listener's position as a staged element of compositional design. It highlights how the mediation of music and media technologies can affect the perceived relationship between the viewer and the performer. It shows how the artist can use these very same music technologies to stage their personas and identities in interesting ways. Most of all, it demonstrates how immersion and interactivity, through dance and gesture, help invigorate interpretations and feelings of embodied experience.

Conclusion

I want to conclude with some final thoughts, pointing towards openings for future research. Concepts of musical immersion and the impacts of immersive and interactive media on popular music production, dissemination, and consumption are part of a burgeoning area for new scholarship. My studies have been primarily concerned with attempting to spur on a discourse within popular musicology that takes more seriously the nebulous nature of multimedia formats and immersive new media, including VR, Dolby Atmos and other 3D sound technologies, and 360º music videos. Moreover, they have demonstrated that there is space within our discipline to expand approaches to music analysis to include mainly the experience of listening and watching music performances.

As a critical study positioned within the field of popular musicology, it opens up for future research, and addresses questions that can be approached in a multitude of ways. First and foremost, I see great potential for further excavating the realm of compositional design and immersive media. In this sense, exploring musicological approaches to immersion and immersive and interactive music media is an ongoing process. Hence, there is a need for more research from both musicologists and practice-based researchers, into the uses of immersive media in live performance practice. In 2019, I had the privilege of working with the Norwegian post-rock band Spurv on a series of 360º concerts in which the band encircled the audience while I and my fellow electronic artist Kristian Isachsen performed 'live remix', sampling the musicians in real time and performing the manipulated sounds back with them during the concert. While this was explicitly an immersive media experience, I would contend that performers not using such technical 360 and 3D setups still regularly engage in creating immersive experiences for concert-goers, including through extending and re-shaping stages, creative sound design, performing in unexpected parts of theaters, and immersive lighting and visual performances.

In my view, it would be relevant for this study to be followed through a variety of approaches. For example, it would be interesting to study the inner workings from studios of up-mixing (the process of re-mixing a stereo recording to a multichannel format) mainstream pop to Dolby Atmos through studio visits, observation, and interviews. Conversely, as probably all of the pop music available on Dolby Atmos format exists both in stereo and 3D, ethnographic audience research with listening tests carried out on samples of the general public would go

a long way to understanding how people re-interpret music in immersive formats, or indeed if they find it more 'immersive' and in what way. There are many research designs in this area which are out of scope or out of the realm of my expertise. For example, in psychoacoustics it is common to evaluate immersive systems using standardized test conditions with validated test signals. However, the laboratory conditions may not always allow for testing with 'real-world' musical examples, and this may lend interesting results, for example attempting to test general perceptions of subject position. Finally, the conceptualization of immersive experience being a result of compositional design implies a conscious choice on the part of pop music artists, recordists, and producers, and ethnographic studies that look specifically to immersion as a compositional element would be interesting and valuable.

Finally, immersion is part and parcel of the pop music experience and hence the pop score. It occurs in any number of contexts, and as I have attempted to demonstrate, it results from both compositional design and interpretive framing. Indeed, what surprised me during this study was that immersive experiences of pop seem to operate in the same way in the more standard formats of stereophonic recordings and music videos. Through the use of space in recorded music and through audiovisual storytelling in music videos, artists, producers, and directors invite us to interpret their work in an infinite number of ways. While this might seem an obvious point of observation, I would argue the contrary. The way that pop music narratives remain so open-ended is a part of this design, and no spatial audio or visual technology can do a better job at enticing us in than the tried-and-true compositional methods they already use to do just that. Still, when these methods are applied within immersive multimedia, it is my experience that the propensity for immersion is enhanced greatly, and for me the most interesting and salient examples of this so far lie in the music videos on Björk's Vulnicura VR album (2019).

Immersion can also be understood as a marketing ploy—co-opted by the music, media, and technology industries to sell us on products and formats that promise to give us these immersive experiences that we crave more readily. As I have demonstrated, it is important to be critical of this ideological (and ultimately capitalist) promise of immersion in pop media. Notwithstanding the ways in which so-called immersive multimedia create the propensity for immersive experience, studying them has offered insights into the properties of immersion in general. For one, they have reconfigured the spatiality of recorded media to center the

viewer/listener on the pop stage. Through this use of spatial audiovisual technology, creators of these musical experiences have given listeners the feeling of a participatory role in the musical performance, and ultimately to the narratives of pop music.

To this end, I have attempted to build on Hawkins' theoretical premises of compositional design in pop music (2002) to expand notions of stylistic and technological codes and to incorporate a theory of immersion. My approach is also indebt to the scholars who have historically identified the recording as central to pop music (Brackett, 2000, 2016; Burns and Lafrance, 2002;Covach, 1999; Frith, 1996; Gracyk, 1996; Hawkins, 1997, 2001, 2002; Moore, 2001, 2012; Moylan, 2002; Tagg, 1987; Théberge, 1997, 2001), where the musical experience itself is always of utmost importance. The question of whether a fallen tree makes any sound is a cliché, and it is not dissimilar to the question of whether music without an audience is in fact music. But this cliché serves to reveal that music, and especially recorded pop music, is always mediated through an audience—there is no way of accessing pop music except qua music. What I have attempted to argue in this thesis is that the way the experience of listening is shaped through the compositional design of audiovisual recordings is integral to the primary pop text. While musical meaning is highly personal and musical interpretations are definitionally subjective, the listening experience itself is a critical component when it comes to understanding compositional design. In other words, the experience of engaging with a pop track or watching a pop video is not separate from the pop score—indeed, it is a critical component of it. Pop music meaning is a dialectic, and the listener is an active participant in the construction of the very media they consume. Reception is thus the cornerstone of interpreting compositional design in the pop score and extracting meaning from pop texts.

This is an established starting point in most pop music analysis and therefore has numerous implications. For one, it implies a greater degree of agency on the part of the listener than is often considered. This is an especially important point in considering mainstream pop, which as many have pointed out is often trivialized in its meaning in sexist, ageist, and racist public discourse that implies the music loved by particular social groups is without value. On the contrary, the very act of listening and interpreting one's favorite music is creating meaning from it— listeners are co-creators of meaning and culture simply in the act of participating in pop music through consumption. In concluding, I'd like to recall a discussion made at the outset with regard the pop score and the pop text, namely that musical

meaning is something that arises in a hermeneutic dialog between the structural elements that make a pop track and the interpretive results of pop music's various texts. This implies that each new listening of a track or viewing of a music video ia about creating a new text; a text that is ultimately the result of both the recorded musical artifact and the exhilarating context of the listening experience.

Article Summaries

As this thesis is article-based, four articles/chapters comprise the second part. The framing chapter has made the argument for considering immersion in audiovisual pop music through the concept of immersive staging, which thematically underpins all these articles as they further explore the perceived relationship between the performer and listener as it is mediated through technology, performativity, compositional design, and aesthetics. The articles, two of which are co-authored with my supervisor, are summarized below:

1. **Immersed in Pop: 3D Music, Subject Positioning, and Compositional Design in The Weeknd's "Blinding Lights" in Dolby Atmos.** In this work, which is published in the autumn 2021 issue of the Journal of Popular Music Studies, I have aimed to address how aesthetic features of pop compositions are altered or maintained in immersive pop music releases, and how different spatial mediums effect compositional design, subject positioning, artists' performativity, and staging. This was done through the invention of a model for immersive music hermeneutics that relates various notions of music technology and production to musicological concepts on performance environment, staging, subject positioning, and compositional design. Finally, the model is demonstrated through a close reading of The Weeknd's 2019 hit 'Blinding Lights', which was released on Dolby Atmos Music.

2. **'A Swarm of Sound': Audiovisual Immersion and Björk's VR video 'Family'** (co-written with Stan Hawkins). The article explores the idea of audiovisual immersion through the portal of the VR (virtual reality) music video. Our focus falls on a close reading Björk's video, 'Family', which addresses questions of immersion in relation to user-experience, staging, and technological innovation. This article draws on the authors' responses to the video by considering the implications of VR immersion in a new generation of music video productions. As part of the methodology on offer, a model for music analysis is devised for conceptualizing virtual audiovisual space (VAVS) and the inextricable relationships between production and compositional design.

3. **Pop Music Diegesis and the 360º Video**. I extend on previous work by asking how immersive pop music video productions shape the narratives that audiovisual pop texts illustrate, which I suggest works through technologically enabled agency and immersion. Taking a hermeneutic approach, I have coined the term pop music diegesis, which helps to explicate the narrative unfolding of a music video in the relationship between the sonic and visual stories. Further, I have considered immersion in 360º videos in the context of pop music diegesis through two modes of interaction and engagement, namely navigational agency and diegetic immersion. Throughout the text, I have supported the theoretical framework with material from the close readings of four 360º music videos available on YouTube: Taryn Southern's Life Support (2018), MUSE's Revolt (2016), The Weeknd's The Hills remix featuring Eminem (2015), and Squarepusher's Stor Eiglass (2015). These videos can be seen on a mobile device in augmented reality (AR) mode, in a VR headset (such as an Oculus Rift), in a head-mounted mobile phone display (such as a Google Cardboard), or simply by mouse navigation on a computer screen.

4. **'Hope to Die': Musicological analysis and queer subjectivity in the music videos of Orville Peck** (co-written with Stan Hawkins and to be published in 2022 in an international collection of essays edited by William Moylan, Lori Burns, and Mike Alleyne). In this chapter we apply a hermeneutic approach couched in analytic methods developed by scholars, such as Hawkins, Moylan, Bresler, Burns, Moore, and others. Examining the music video 'Hope to Die' by queer country icon, Orville Peck, we attempt to unravel the sonic details of production within compositional design while making a case for audiovisual representation. While this work is a break from the preceding articles that consider so-called 'immersive' music recordings and videos, it seeks to show that concepts of immersion and interactivity are relatable not only to immersive and interactive media, but to music recordings and videos in general. Our work asks questions about the staging of gender and sexual identity, and how immersion can operate with relation to identity representations and aesthetic endeavors.

References

Auslander, P. (2008). *Liveness: Performance in a Mediatized Culture* (2nd ed.). New York: Routledge.

Auslander, P. (2009). Musical Persona: The Physical Performance of Popular Music. In D. B. Scott (Ed.), *The Ashgate Research Companion to Popular Musicology* (pp. 303-315). Surrey, UK: Ashgate.

Auslander, P. (2021). *In Concert: Performing Musical Persona*. Ann Arbor, MI: University of Michigan Press.

Auslander, P., & Inglis, I. (2016). "Nothing is Real": The Beatles as Virtual Performers. In S. Whiteley & S. Rambarran (Eds.), *The Oxford Handbook of Music and Virtuality* (pp. 35-51). New York: Oxford University Press.

Björnberg, A. (2009). Learning to Listen to Perfect Sound: Hi-Fi Culture and Changes in Modes of Listening, 1950-80. In D. B. Scott (Ed.), *The Ashgate Research Companion to Popular Musicology* (pp. 105-129). Surrey, UK: Ashgate.

Blum, S. (1993). In Defence of Close Reading and Close Listening. *Current Musicology, 53*, 41-54.

Brackett, D. (2000). *Interpreting Popular Music* (2nd ed.). Berkeley: University of California Press.

Brackett, D. (2016). *Categorizing sound: genre and twentieth-century popular music*. Berkeley: University of California Press.

Brøvig-Hanssen, R., & Danielsen, A. (2016). *Digital Signatures: The Impact of Digitization on Popular Music Sound*. Cambridge: MIT Press.

Burns, L. (2016). The Concept Album as Visual-Sonic-Textual Spectacle: The Transmedial Storyworld of Coldplay's Mylo Xyloto. *IASPM Journal, 6*(2), 91-116.

Burns, L. (2018). Interpreting Transmedia and Multimodal Narratives: Steven Wilson's "The Raven That Refused to Sing". In C. Scotto, K. Smith, & J. Brackett (Eds.), *The Routledge Companion to Popular Music Analysis: Expanding Approaches* (pp. 95-113). New York: Routledge.

Burns, L., & Hawkins, S. (2019a). Introduction. In L. Burns & S. Hawkins (Eds.), *The Bloomsbury Handbook of Popular Music Video Analysis* (pp. 1-9). New York: Bloomsbury.

Burns, L., & Hawkins, S. (Eds.). (2019b). *The Bloomsbury Handbook of Popular Music Video Analysis*. New York: Bloomsbury.

Burns, L., & Lafrance, M. (2017). Gender, Sexuality, and the Politics of Looking in Beyoncé's 'Video Phone' (Featuring Lady Gaga). In S. Hawkins (Ed.), *The Routledge Research Companion to Popular Music and Gender* (pp. 102-116). New York: Routledge.

Burns, L., Lafrance, M., & Hawley, L. (2008). Embodied Subjectivities in the Lyrical and Musical Expression of PJ Harvey and Björk. *Music Theory Online, 14*(4). Retrieved from https://mtosmt.org/issues/mto.08.14.4/mto.08.14.4.burns_lafrance_hawley.html

Burns, L., & Watson, J. (2010). Subjective Perspectives through Word, Image and Sound: Temporality, narrative agency and embodiment in the Dixie

Chicks' video 'Top of the World. *Music, Sound, and the Moving Image, 4*(1), 3-37.

Burns, L., & Woods, A. (2018). Rap Gods and Monsters: Words, Music, and Images in the Hip-Hop Intertexts of Eminem, Jay-Z, and Kanye West. In L. Burns & S. Lacasse (Eds.), *The Pop Palimpsest: Intertextuality in Recorded Popular Music* (pp. 215-251). Ann Arbor, MI: University of Michigan Press.

Butler, J. (1993). *Bodies that Matter: On the Discursive Limits of "Sex"*. New York: Routledge.

Camilleri, L. (2010). Shaping sounds, shaping spaces. *Popular Music, 29*(2), 199-211.

Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. New York: Oxford University Press.

Collins, K. (2007). Video Games Killed the Cinema Star: It's Time for a Change in Studies of Music and the Moving Image. *Music, Sound, and the Moving Image, 1*(1), 15-19.

Collins, K. (2013). *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. Cambridge: MIT Press.

Collins, K., & Dockwray, R. (2015). Sonic Proxemics and the Art of Persuasion: An Analytical Framework. *Leonardo Music Journal, 25*, 53-56.

Covach, J. (1999). Popular Music, Unpopular Musicology. In N. Cook & M. Everist (Eds.), *Rethinking Music* (pp. 452-470). Oxford: Oxford University Press.

Csikszentmihalyi, M. (1990). *Flow. The Psychology of Optimal Experience*. New York: Harper Perennial.

Danielsen, A. (2006). *Presence and pleasure: the funk grooves of James Brown and Parliament*. Middletown, CT: Wesleyan University Press.

Danielsen, A. (2015). Metrical Ambiguity or Microrhythmic Flexibility? Analysing Groove in 'Nasty Girl' by Destiny's Child. In R. Von Appen, A. Doehring, D. Helms, & A. F. Moore (Eds.), *Song Interpretation in 21st-Century Pop Music* (pp. 53-72). Surrey: Ashgate.

Danielsen, A., & Hawkins, S. (2020). "The Right Amount of Odd": Vocal Compulsion, Structure, and Groove in Two Love Songs from Around the World in a Day. *Popular Music and Society, 43*(3), 1-19. doi:10.1080/03007766.2020.1757814

DeFrantz, T. F. (2004). The Black Beat Made Visible: Hip Hop Dance and Body Power. In A. Lepecki (Ed.), *Of the Presence of the Body: Essays on Dance and Performance Theory* (pp. 64-81). Middletown, CT: Wesleyan University Press.

DeNora, T. (2000). *Music in everyday life*. Cambridge: Cambridge University Press.

Dibben, N. (2012). The Intimate Singing Voice: Auditory Spatial Perception and Emotion in Pop Recordings. In D. Zakharine & N. Meise (Eds.), *Electrified Voices: Medial, Socio-Historical and Cultural Aspects of Voice Transfer* (pp. 107-122). Göttingen, DE: V&R unipress.

Dibben, N. (2013). Visualizing the App Album with Björk's Biophilia. In C. Vernallis, J. Richardson, & A. Herzog (Eds.), *The Oxford Handbook of Sound and Image in Digital Media* (pp. 682-704). New York: Oxford University Press.

Dickinson, K. (2004). 'Believe': vocoders, digital female identity, and camp. In S. Whiteley, A. Bennett, & S. Hawkins (Eds.), *Music, Space and Place* (pp. 163-179). Aldershot, UK: Ashgate.

Dockwray, R., & Moore, A. F. (2010). Configuring the sound-box 1965–1972. *Popular Music, 29*(2), 181-197.

Douglas, Y., & Hargadon, A. (2000). *The pleasure principle: immersion, engagement, flow*. Paper presented at the 11th ACM on Hypertext and Hypermedia, San Antonio, TX.

Eidsheim, N. S. (2015). *Sensing Sound: Singing and Listening as Vibrational Practice*. Durham, NC: Duke University Press.

Eidsheim, N. S. (2019). *The Race of Sound: Listening, Timbre, and Vocality in African American Music*. Durham, NC: Duke University Press.

Eidsheim, N. S., & Meizel, K. (Eds.). (2019). *The Oxford Handbook of Voice Studies.* Oxford: Oxford University Press.

Fathallah, J. (2021). Is stage-gay queerbaiting? The politics of performative homoeroticism in emo bands. *Journal of Popular Music Studies, 33*(1), 121-136.

Frith, S. (1996). *Performing Rites: On the Value of Popular Music*. Cambridge, MA: Harvard University Press.

Frith, S., & Zagorski-Thomas, S. (2012). Introduction. In S. Frith & S. Zagorski-Thomas (Eds.), *The Art of Record Production: An Introductory Reader for a New Academic Field* (pp. 1-9). Surrey, UK: Ashgate.

Gibson, D. (1997). *The Art of Mixing: A Visual Guide to Recording, Engineering, and Production*. Vallejo, CA: Mix Books.

Gibson, J. J. (1977). The Theory of Affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting and Knowing: Toward and Ecological Psycology*. Mahwah, NJ: Lawrence Erlbaum.

Gibson, J. J. (2015). *The Ecological Approach to Visual Perception* (3rd ed.). New York: Psychology Press.

Gorbman, C. (1980). Narrative Film Music. *Yale French Studies, 60*, 183-203. doi:10.2307/2930011

Gracyk, T. (1996). *Rhythm and Noise: An Aesthetics of Rock*. Durham, NC: Duke University Press.

Greengard, S. (2019). *Virtual Reality*. Cambridge: MIT Press.

Hansen, K. A. (2017a). *Fashioning Pop Personae: Gender, Personal Narrativity, and Converging Media in 21st Century Pop Music.* (Ph.D). University of Oslo, Norway.

Hansen, K. A. (2017b). Holding on for dear life: Gender, celebrituy status, and vulnerability-on-display in Sia's 'Chandelier'. In S. Hawkins (Ed.), *The Routledge Reserach Companion to Popular Music and Gender* (pp. 89-101). New York: Routledge.

Hansen, K. A. (2019). (Re)Reading Pop Personae: A Transmedial Approach to Studying the Multiple Construction of Artist Identities. *Twentieth-Century Music, 16*(3), 501-529. doi:10.1017/S1478572219000276

Hansen, K. A., Askerøi, E., & Jarman, F. (2021a). Introduction: a musicology of popular music and identity. In K. A. Hansen, E. Askerøi, & F. Jarman (Eds.), *Popular Music and Identity: Essays in Honour of Stan Hawkins*. New York: Routledge.

Hansen, K. A., Askerøi, E., & Jarman, F. (Eds.). (2021b). *Popular Musicology and Identity: Essays in Honor of Stan Hawkins*. New York: Routledge.

Hawkins, S. (1992). Prince: harmonic analysis of 'Anna Stesia'. *Popular Music, 11*(3), 325-335.

Hawkins, S. (1997). The Pet Shop Boys: Musicology, masculinity and banality. In S. Whiteley (Ed.), *Sexing the Groove*. London: Routledge.

Hawkins, S. (2001). Musicological Quagmires in Popular Music: Seeds of Detailed Conflict. *Popular Music Online*. Retrieved from http://www.popular-musicology-online.com/issues/01/hawkins.html

Hawkins, S. (2002). *Settling the pop score: Pop texts and identity politics*. Burlington, VT: Ashgate.

Hawkins, S. (2003). Feel the beat come down: house music as rhetoric. In A. F. Moore (Ed.), *Analyzing Popular Music* (pp. 80-102). Cambridge, UK: Cambridge University Press.

Hawkins, S. (2004). On performativity and production in Madonna's 'Music'. In S. Whitely, A. Bennett, & S. Hawkins (Eds.), *Music, Space and Place: Popular Music and Cultural Identity*. Surrey, UK: Ashgate.

Hawkins, S. (2008). Temporal Turntables: On Temporality and Corporeality in Dance Culture. In S. Baur, J. Warwick, & R. Knapp (Eds.), *Musicological Identities: Essays in Honor of Susan McClary* (pp. 121-134). New York: Routledge.

Hawkins, S. (2009). *The British pop dandy: masculinity, popular music and culture*. New York: Routledge.

Hawkins, S. (2012). 'Great, Scott!'. In S. Hawkins (Ed.), *Critical Musicological Reflections: Essays in Honour of Derek B. Scott*. Surrey, UK: Ashgate.

Hawkins, S. (2013). Aesthetics and Hyperembodiment in Pop Videos: Rihanna's "Umbrella". In J. Richardson, C. Gorbman, & C. Vernallis (Eds.), *The Oxford Handbook of New Audiovisual Aesthetics* (pp. 466-482). Oxford: Oxford University Press.

Hawkins, S. (2016). *Queerness in Pop Music: Aesthetics, Gender Norms, and Temporality*. New York: Routledge.

Hawkins, S. (2018). Performative Strategies and Musical Markers inthe Eurythmics' "I Need a Man". In L. Burns & S. Lacasse (Eds.), *The Pop Palimpsest: Intertextuality in Recorded Popular Music* (pp. 252-270). Ann Arbor, MI: Univeristy of Michigan Press.

Hawkins, S. (2020). Personas in Rock: "We Will, We Will Rock You". In A. F. Moore & P. Carr (Eds.), *The Bloomsbury Handbook of Rock Music Research* (pp. 239-254). London: Bloomsbury.

Hawkins, S. (Ed.) (2017). *The Routledge Research Companion to Popular Music and Gender*. New York: Routledge.

Hebrank, J., & Wright, D. (1974). Spectral cues used in the localization of sound sources on the median plane. *Journal of the Acoustical Society of America, 56*. doi:10.1121/1.1903520

Jirsa, T., & Korsgaard, M. B. (2019). The Music Video in Transformation: Notes on a Hybrid Audiovisual Configuration. *Music, Sound, and the Moving Image, 13*(2), 111-122.

Johnston, S. (1999). Structuralism and its Aftermath. In P. Cook & M. Bernink (Eds.), *The Cinema Book* (2nd ed., pp. 323-341). London: British Film Institute.

Kassabian, A. (2013). The end of diegesis as we know it? In J. Richardson, C. Gorbman, & C. Vernallis (Eds.), *The Oxford Handbook of New Audiovisual Aesthetics*. Oxford: Oxford University Press.

Kassabian, A. (2017). "You mean I can make a TV show?": Web series, assertive music, and African American women producers. In S. Hawkins (Ed.), *The Routledge Research Companion to Popular Music and Gender* (pp. 79-88). London: Routledge.

Korsgaard, M. B. (2017). *Music Video After MTV: Audiovisual Studies, New Media, and Popular Music*. New York: Routledge.

Korsgaard, M. B. (2019). SOPHIE's 'Faceshopping' as (Anti-)Lyric Video. *Music, Sound, and the Moving Image, 13*(2), 209-230.

Kramer, L. (1993). Music Criticism and the Postmodernist Turn: In Contrary Motion with Gary Tomlinson. *Current Musicology, 53*, 25-35.

Kramer, L. (2011). *Interpreting Music*. Berkeley: University of California Press.

Kraugerud, E. (2021). *Come Closer: Acousmatic Intimacy in Popular Music Sound.* (PhD). University of Oslo,

Lacasse, S. (2000a). Intertextuality and Hypertextuality in Recorded Popular Music. In M. Talbot (Ed.), *The Musical Work: Reality or Invention?* (pp. 35-58). Liverpool: Liverpool University Press.

Lacasse, S. (2000b). *'Listen to my voice': the evocative power of vocal staging in recorded rock music and other forms of vocal expression.* (PhD). University of Liverpool, UK.

LaFrance, M. (2013). Celebrity, Spectacle, and Surveillance: Understanding Lady Gaga's 'Paparazzi' and 'Telephone' through Music, Image, and Movement In M. Iddon & M. L. Marshall (Eds.), *Lady Gaga and Popular Music*. New York: Routledge.

Liljedahl, A. A. (2019). Musical Pathfinding; or How to Listen to Interactive Music Video. *Music, Sound, and the Moving Image, 13*(2), 165-185. doi:https://doi.org/10.3828/msmi.2019.10

McClary, S. (1991). *Feminine endings: Music, gender, and sexuality*. Minneapolis: University of Minnesota Press.

McClary, S. (1993). Reshaping a Discipline: Musicology and Feminism in the 1990s. *Feminist Studies, 19*(2), 399-423.

McIntyre, P. (2012). *Creativity and Cultural Productino: Issues for Media Practice*. New York: Palgrave Macmillan.

McLeod, K. (2016). Living in the Immaterial World: Holograms and Spirituality in Recent Popular Music. *Popular Music and Society, 39*(5), 501-515. doi:10.1080/03007766.2015.1065624

Middleton, R. (1990). *Studying popular music*. Buckingham, UK: Open University Press.

Middleton, R. (2000). Introduction. In R. Middleton (Ed.), *Reading Pop: Approaches to Textual Analysis in Popular Music* (pp. 1-19). Oxford: Oxford University Press.

Miles, C. (2020). Black Rural Feminist Trap: Stylized and Gendered Performativity in Trap Music. In *Journal of Hip Hop Studies* (Vol. 7, pp. 44-70).

Moore, A. F. (2001). *Rock: The Primary Text; Developing a Musicology of Rock* (2nd ed.). Surrey: Ashgate.

Moore, A. F. (2002). Authenticity as authentication. *Popular Music, 21*(2), 209-223.

Moore, A. F. (2003). Introduction. In A. F. Moore (Ed.), *Analyzing Popular Music* (pp. 1-15). Cambridge: Cambridge University Press.

Moore, A. F. (2012). *Song Means: Analysing and Interpreting Recorded Popular Song*. Surrey: Ashgate.

Moore, A. F. (2013). An Interrogative Hermeneutics of Popular Song. *El Oído Pensante, 1*, 7-27.

Moore, A. F., & Dockwray, R. (2008). The establishment of the virtual performance space in rock. *Twentieth-Century Music, 5*(2), 219-241.

Moore, A. F., Schmidt, P., & Dockwray, R. (2009). A hermeneutics of spatialization for recorded song. *Twentieth-Century Music, 6*, 83-114.

Morie, J. F. (2007). Performing in (virtual) spaces: Embodiment and being in virtual environments. *International Journal of Performance Arts and Digital Media, 3*, 123-138. doi:10.1386/padm.3.2-3.123_1

Moylan, W. (2002). *The Art of Recording: Understanding and Crafting the Mix*. New York: Focal Press.

Moylan, W. (2012). Considering space in recorded music. In S. Frith & S. Zagorski-Thomas (Eds.), *The Art of Record Production: An Introductory Reader for a New Academic Field* (pp. 163-188). Surrey: Ashgate.

Moylan, W. (2020). *Recording Analysis: How the Record Shapes the Song*. New York: Routledge.

Negus, K. (1999). *Music Genres and Corporate Cultures*. London: Routledge.

Negus, K., & Pickering, M. (2004). *Creativity, Communication and Cultural Value*. Londone: Sage.

Parsons, A. (1975). Four Sides of the Moon. *Studio Sound*.

Perrott, L. (2019). 'Accented' Music Video: Animating Memories of Migration in 'Rocket Man'. *Music, Sound, and the Moving Image, 13*(2), 123-146.

Povey, G. (2016). *The Complete Pink Floyd: The Ultimate Reference*. New York: Sterling.

Rambarran, S. (2021). *Virtual Music: Sound, Music, and Image in the Digital Era*. New York: Bloomsbury Academic.

Richardson, J., & Gorbman, C. (2013). Introduction. In J. Richardson, C. Gorbman, & C. Vernallis (Eds.), *The Oxford Handbook of New Audiovisual Aesthetics* (pp. 3-35). Oxford: Oxford University Press.

Richardson, K. (2003). Another Phase of the Moon. *Sound & Vision.* Retrieved from https://www.soundandvision.com/content/another-phase-moon

Roffler, S. K., & Butler, R. A. (1968). Localization of Tonal Stimuli in the Vertical Plane. *The Journal of the Acoustical Society of America.* doi:10.1121/1.1910977

Rose, T. (2008). *The Hip Hop Wars: What We Talk About When We Talk About Hip Hop—and Why It Matters*. New York: Basic Books.

Sandve, B. (2014). Staging the Real: Identity politics and urban space in mainstream Norwegian rap music. In.

Scott, D. B. (1990). Music and Sociology for the 1990s: A Changing Critical Perspective. *The Musical Quarterly, 74*(3), 385-410.

Scott, D. B. (2009). Introduction. In D. B. Scott (Ed.), *The Ashgate Research Companion to Popular Musicology* (pp. 1-21). Surrey, UK: Ashgate.

Senior, M. (2012). *Mixing Secrets*. New York: Focal Press.

Simon Frith, S. Z.-T. (2012). The Art of Record Production: An Introductory Reader for a New Academic Field. In: Ashgate Pub Co.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised Sound, 2*(2), 107-126.

Smalley, D. (2007). Space-form and the acousmatic image. *Organised Sound, 12*(1), 35-58.

Stefani, G., & Fiori, U. (1984). An Interview with Gino Stefani. *IASPM Newsletter, 5*, 18-19.

Strachan, R. (2017). *Sonic Technologies: Popular Music, Digital Culture and the Creative Process*. New York: Bloomsbury.

Street, J. (2011). *Music and Politics*. New York: Wiley.

Tagg, P. (1979). *Kojak: Fifty Seconds of Television Music.* (PhD). University of Göteborg, Gothenburg, SE.

Tagg, P. (1982). Analysing popular music: theory, method and practice. In *Popular Music* (Vol. 2, pp. 37-67).

Tagg, P. (1987). Musicology and the semiotics of popular music. *Semiotica, 66*, 279-298.

Tannen, D. (1993). What's in a Frame?: Surface Evidence for Underlying Expectations. In D. Tannen (Ed.), *Framing in Discourse* (pp. 14-56). New York: Oxford University Press.

Théberge, P. (1997). *Any sound you can imagine: making music/consuming technology*. Hanover, N.H: Wesleyan University Press.

Théberge, P. (2001). 'Plugged In': Technology and Popular Music. In S. Frith, W. Straw, & J. Street (Eds.), *Cambridge Companion to Pop and Rock* (pp. 3-25). Cambridge: Cambridge University Press.

Thompson, P. (2018). Creativity in the Recording Studio: Alternative Takes. In K. Spracklen & K. Fox (Eds.), *Leisure Studies in a Global Era*. Cham, Switzerland: Springer.

Thompson, P., & McIntyre, P. (2013). Rethinking Creative Practice In Record Production and Studio Recording Education: Addressing The Field. *Journal on the Art of Record Production*(8). Retrieved from http://www.arpjournal.com/asarpwp/rethinking-creative-practice-in-record-production-and-studio-recording-education-addressing-the-field/

Tomlinson, G. (1993). Musical Pasts and Postmodern Musicologies: A Response to Lawrence Kramer. *Current Musicology, 53*, 18-24.

Vernallis, C. (2004). *Experiencing Music Video: Aesthetics and Cultural Context*. New York: Columbia University Press.

Vernallis, C. (2008). Music video, songs, sound: experience, technique and emotion in Eternal Sunshine of the Spotless Mind. *Screen, 49*(3), 277-297.

Vernallis, C., Herzog, A., & Richardson, J. (Eds.). (2013). *The Oxford Handbook of Sound and Image in Digital Media*. Oxford: Oxford University Press.

Vernallis, C., & Ueno, H. (2013). Interview with Music Video Director and Auteur Floria Sigismondi. *Music, Sound, and the Moving Image, 7*(2), 167-194.

Wallis, R., & Lee, H. (2015). The Effect of Interchannel Time Difference on Localisation in Vertical Stereophonty. *Journal of the Audio Engineering Society, 63*(10), 767-776. doi:10.17743/jaes.2015.0069

Walser, R. (1993). *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*. Middletown, CT: Wesleyan University Press.

Walther-Hansen, M. (2015). Sound Events, Spatiality and Diegesis – The Creation of Sonic Narratives in Music Productions. *Danish Musicology Online*, 29-46.

Whiteley, S. (2000). *Women and Popular Music: Sexuality, Identity and Subjectivity*. London: Routledge.

Whiteley, S. (2016). Introduction. In S. Whiteley & S. Rambarran (Eds.), *The Oxford Handbook of Music and Virtuality* (pp. 1-10). New York: Oxford University Press.

Whiteley, S. (Ed.) (1997). *Sexing the Groove: Popular Music and Gender*. London: Routledge.

Whiteley, S., Bennett, A., & Hawkins, S. (2004). Introduction. In S. Whiteley, A. Bennett, & S. Hawkins (Eds.), *Music, Space and Place: Popular Music and Cultural Identity* (pp. 1-22). Aldershot, UK: Ashgate.

Whiteley, S., & Rambarran, S. (Eds.). (2016). *The Oxford Handbook of Music and Virtuality*. Oxford: Oxford University Press.

Wicke, P. (2009). The Art of Phonography: Sound, Technology and Music. In D. B. Scott (Ed.), *The Ashgate Research Companion to Popular Musicology* (pp. 147-168). Surrey, UK: Ashgate.

Winters, B. (2010). The non-diegetic fallacy: Film, music, and narrative space. *Music and Letters, 91*(2), 224-244. doi:10.1093/ml/gcq019

Zagorski-Thomas, S. (2010). The stadium in your bedroom: functional staging, authenticity and the audience-led aesthetic in record production. *Popular Music, 29*(2), 251-266.

Zagorski-Thomas, S. (2014). *The Musicology of Record Production*: Cambridge University Press.

Article 1 – Immersed in Pop: 3D Music, Subject Positioning, and Compositional Design in The Weeknd's "Blinding Lights for Dolby Atmos

Zack Bresler

Published in the Journal of Popular Music Studies, 33(3), September 2021

Introduction

While stereophonic sound has been the dominant release format for popular music for decades, innovation into audio formats has persisted outside the pop sphere, and sometimes attempts are made to bridge such innovations with popular music and culture. In the contemporary multimedia landscape, this includes technologies such as virtual reality[1] and so-called 'immersive' formats[2] like Dolby Atmos and Sony 360 Reality Audio, both of which are 3D sound formats which began implementation into the streaming services Amazon Prime Music HD, Deezer HiFi, and Tidal HiFi in late 2019 and early 2020.[3] However, there is a persistent notion among creators and scholars of popular music that stereo sound is somehow a defining feature of pop music[4]—that at some level, be it functional, economic, or aesthetic, stereo is the de facto frame for the pop stage. While it is difficult to argue against the fact that stereophonic sound is central to popular music production practices, the notion of its default status is challenged through the ever-increasing use of immersive and interactive media technologies on streaming

---

[1] For example, in autumn 2019, Björk released an album of music videos in Virtual Reality entitled *Vulnicura VR*.

[2] The terms 'immersive audio' and 'immersive format' seem at present to be the standard terms used in the music technology field to describe any multichannel audio format which is at least '2.5D', or hemispheric sound either over loudspeakers (surround sound with height) or in binaural over headphones (as is typical in virtual and augmented reality). In 2019, the Audio Engineering Society held the Immersive and Interactive Audio conference, which brought together academics and industry partners to "explore the unique space where interactive technologies and immersive audio meet and aims to exploit the synergies between these fields" (http://www.aes.org/conferences/2019/immersive/).

[3] https://www.digitaltrends.com/home-theater/what-is-dolby-atmos-music-and-how-to-get-it/

[4] In the opening of their anthology on multichannel audio, Théberge, Devine and Everett claim that "stereo is a living part of sound culture" (Théberge, Devine, and Everrett 2015, 1). While the research in this volume is of value, it also is at the centre of a romanticised narrative that puts stereophonic sound at the end of music recording's inevitable progression through technology. This narrative seems to be one somewhat shaky ground given the rapid emergence of 'new media' technologies, as described above, that challenge stereo's predominance in all forms of media, including popular music.

services such as Tidal and Spotify, and social media platforms like Facebook and YouTube.

How are the aesthetics of pop compositions altered or maintained in immersive music productions? How does this effect compositional design, performativity, staging, and space? This article attempts to address the changing effects immersive and interactive technologies have on these aspects of popular music by suggesting a model for close analysis of such music. This model will help the reader better understand immersive popular music by demonstrating how music production technologies and practices relate to already established ideas about music interpretation and the relationship of the artist and the viewer of a pop composition. In an effort to demonstrate the model's efficacy, I turn to a discussion of a song by the R&B artist The Weeknd entitled 'Blinding Lights', which was mixed in both stereo and Dolby Atmos 3D formats and released in late 2019.

Dolby Atmos is a flexible, object-based 3D audio format. In short, this means that the format is based around a standard surround sound configuration (such as the 5.1 system common to home theater systems), with an added processor for handling sound objects, which can be in any location in 3D space and rendered at playback to the user's sound system. For example, a user with a 5.1.2 Dolby Atmos system has five speakers in surround sound, one subwoofer, and two speakers elevated above their main left and right speakers. In 2019, Dolby announced "Atmos Music," which promised to deliver thousands of audio-only music releases on various streaming services in the format in the coming few years. Notably, many of The Weeknd's most popular releases, as well as those of pop artists from Lizzo to Elton John, are currently available in Dolby Atmos on a growing number of platforms.

Modelling immersive popular music

For some time, popular musicologists have suggested approaches that aid and assist interpretations, recognizing that we are nonetheless listeners, fans, and participants in popular music and popular culture (Hansen et al., 2021; Scott, 2009). Accordingly, my approach to pop music analysis centers primarily around the identification of musical codes.[5] Therefore, the question of who is listening

---

[5] By 'musical codes', primarily I am referring to semiotic and hermeneutic approaches to music analysis (i.e. close readings), which consider the music as 'text' which is interpreted through the identification of various features (technological, aesthetic, cultural, functional, etc.) which are referred to as codes. For a

becomes particularly relevant, acknowledging that my interpretation will surely differ from the reader's as my background, tastes, location, time, etc., will lead me to identify some codes as significant and others as irrelevant. David Brackett reminds us that codes are never decoupled from their interpreters, and that "listener 'competence'… refers to the range of subject positions available to a listener dependent on that individual's history and memory" (Brackett, 2000, p. 13). Similarly, Stan Hawkins has problematized the identification of musical codes (Hawkins, 2002), insisting that "there is always a sense of legitimacy in one's own brand of hermeneutics that seeks to validate the means of one's craft"(Hawkins, 2001). Given that pop texts generate a range of possible subject positions from which to interpret, it follows that they can be understood as staged. In other words, each interpretive subject positioning represents a staging of the current listener that is as nuanced and complex as the listener's competence allows.

Central to the analysis is the concept of staging, which requires some unpacking. Competing with each other are two related but different ways of thinking about staging. In one sense, staging refers to the physical or perceptual positioning of sound objects in the recorded space—the placement of the performance on the stage. Although they do not explicitly employ the term 'staging', this definition of the term is highly congruent with Allan Moore's soundbox (Moore, 2001, 2012) and William Moylan's perceived performance environment models (Moylan, 2002, 2012). Of course, these models, like this work and that of many others, use this concept to move from the sonic perceptual towards the metaphorical and musicological, considering how the artist constructs aspects of performance and identity that contain deeper meaning for listeners. For example, Moore bridges the perceptual with the hermeneutic by introducing proxemic relationships as ways to interpret aspects of the performance persona (Moore, 2012, pp. 185–186). Philip Auslander discusses the ways in which liveness both constructs and is constructed by recorded music (Auslander, 2008, pp. 73–127), and I would suggest that his argument is about how the staging of rock records dictates the staging of live performances and vice versa, which of course has a huge impact on the listener's interpretations of meaning, subject position, persona, and authenticity. This article uses concepts of staging in similar

more in-depth discussion, see the introduction to Stan Hawkins' book Settling the Pop Score (2002), which deals with musical codes and hermeneutic methodologies extensively.

ways to these to show how 3D music reconfigures the performance stage, both in terms of its perceived physical parameters and its hermeneutic relationship to listeners.

In focusing my analytic intentions, I have devised a model that serves to deconstruct various aspects of interpretation to relate particular observations about music production aesthetic features to musicological discourses. This model has three areas of analysis, which are *balance and proxemic distance*, *performativity and vocal staging*, and *subject positioning and perception* (fig. 1).
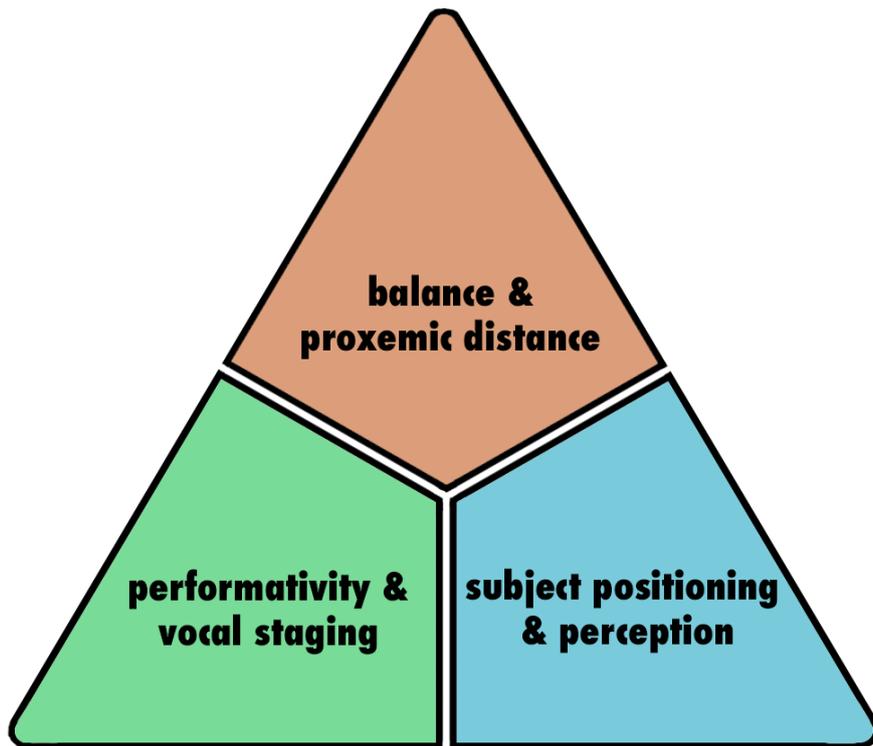
Figure 1: Model for hermeneutic analysis of 3D music

Balance and proxemic distance

The technological means by which panning and balance are achieved in pop mixes are important to understand as they are part of a large palate of tools that create a stage for performers to position themselves in a variety of proxemic contexts (Collins & Dockwray, 2015). Allan Moore refers to Hall's model of proxemics (See: Hall, 1966), or social distances, in reference to popular music spatiality to describe both the perceived physical distance of the performer to the listener as well as "the degree of congruence between a persona and the personic

environment" (Moore, 2012, p. 186). Personae[6] are staged hermeneutically by listeners as they interpret their relationship to performers, and this relationship is both literally and metaphorically spatial. For example, an artist may choose to use very little reverb while singing very quietly into the microphone, such as the famous opening to Salt-N-Pepa's 'Push It', which suggests an intimate and highly sexualized proxemic relation to the listener. In the same song, when the voices of Salt-N-Pepa are heard in rap/hype contexts, the staging is more distant—there is a perceptible amount of reverb and delay, and the rappers are clearly vocalizing at a louder volume, suggesting a social distance that is interpreted as a stage address to a crowd.

Given that the primary mode of dissemination for popular music has been stereophonic sound, it makes sense to begin examining this spatial relationship through a discussion of compositional staging norms in stereo music. In particular, the normative structure of the spatial placement, or balance, of instruments, voices, sounds, and effects in pop music mixes is considered here. The balance of a pop mix has been considered in a number of ways, but here I begin with the idea of the 'diagonal mix', a term coined by Moore that describes a mixing structure that emerged shortly after the introduction of stereo sound where "a lead vocal, a snare drum, and the harmonic bass… are situated centrally on a (very) slight diagonal" (Moore, 2012, p. 32). It is clear that the panning of lead elements in the center while balancing secondary elements across the stereo image is a practice carried on through contemporary popular music production, as evidenced by the emphasis placed on this kind of balance by authors of modern mixing method books. David Gibson's influential manual *The Art of Mixing* refers to the panning of important elements into the center of the stereo image as a presupposition: "As you probably have noticed in mixes, some sounds are right out in front (normally vocals and lead instruments)" (Gibson, 1997, p. 10). Going a bit further, mixing engineer Mike Senior argues for this panning structure from a more practical perspective, suggesting that the reason we put these elements "in the middle of the stereo image

---

[6] The term 'persona' in this passage requires unpacking, as it is not used consistently across literature. Here, Moore is referring specifically only to the constructed performer as they exist in the musical text, avoiding extra-musical factors that others explicitly consider as part of the construction of personae. For example, Philip Auslander is careful to distinguish between the 'real person', 'performance persona', and 'character' (2009, 305), while Moore openly conflates the boundaries between these notions (2012, 180-181) and avoids consideration of factors outside the musical text. For a detailed discussion of these definition, see Hansen 2017. (Auslander, 2009; Hansen, 2017; Moore, 2012)

[is] because they'll be the ones that survive best in mono" (Senior, 2012, pp. 126–127). Considering the secondary sonic elements, we get a clearer and more complete picture of balance structure in pop mixes. Gibson claims that "for some instruments, the traditions for the specific placement of left to right have become very strictly enforced" (Gibson, 1997, p. 99). Senior gives a method for accomplishing this balance, which he calls "opposition panning," in which balance is created through panning sources opposite to one another based on their "musical function," such that anything panned to the left, for example, should have a musical equivalent on the right (Senior, 2012, p. 127).

In general, popular music productions in surround and 3D formats have followed a similar panning and balance pattern, referred to as the *front stage image* (Gerzon, 1992; Glasgal, 2001; Moylan, 2002). Similar to the diagonal mix, this spatial structure recreates the stereo sound-stage metaphor in front of the listener, using additional side, rear, and height dimensions for widening stereo images, modelling acoustic spatialities such as performance halls and recording studios, and special effects such as placing the background vocals in the literal background. An iconic example of special effects in surround popular music comes from Allan Parsons' famous quadraphonic mixes of Pink Floyd's 'Money' from *Dark Side of the Moon*, where, in the introduction, the listener is immersed in a cacophony of ringing cash registers and dropping coins. The scene is seemingly without a front image for a few seconds, panning wildly around the listener until the famous 7/4 bass line comes in front and center to stabilize and establish the stage dimensions (Pink Floyd, 1973).

One of the most important mixing tools that impacts on proxemic relationships is reverb and delay which often combine with panning in stereo to create the illusions of depth and distance. This works because, without visual cues, we tend to relate sound sources to imagined causes, which Denis Smalley refers to as "source bonding" (Smalley, 1997, p. 110). Moore's 'sound-box' model describes stereo depth as a function of relative volume and relative reverberation (Moore, 2012, pp. 30–31), as does William Moylan's 'perceived performance environment' model, which has six characteristics, four of which deal with relative relationships of various aspects of reverberation and delay (Moylan, 2012, p. 164). Since real-world sounds propagate our environments full of reflections, they are heard with echoes and natural reverberations, which can of course be replicated with music processing equipment and software. Of course, while realist representations of acoustics may be a goal of some music, popular music often mixes and matches

different spatialities for artistic effect. As Ragnhild Brøvig-Hanssen and Anne Danielsen point out, when hearing multiple spatialities simultaneously "we do not draw upon any given experience with a particular space but are rather forced to attempt an awkward synthesis of a number of such spaces" (Brøvig-Hanssen & Danielsen, 2016, p. 33). Such "surrealism" of spatiality need not be 'unnatural', as typifies popular music as much as the sense of a normal listening process (Brøvig-Hanssen, 2013, pp. 14–22).

Staging in immersive pop mixes can have a great impact on perceptions of proxemic distance. For one, the front stage image in surround and 3D music means that the mix can contain spatio-acoustic detail of much higher resolution, in some cases creating the illusion of replacing the reverberation of the listening space with that of the virtual recorded space. This added acoustics footprint comes at a price, which is that, if used globally, the additional reverb can have the effect of distancing the mix from the listener significantly. The potential downside is that the physical distancing effects the perceived proxemic relation between performer and listener, which can greatly affect the perceived meaning for the listener. In the example I will look at later, a different approach was taken, in which certain elements were given greater spatio-acoustic detail (i.e. reverb and delay) while others were not, allowing those less-processed sounding elements to be interpreted as being closer. It is clear that immersive spatial configurations, for both technical and artistic reasons, can have a great effect on perceptions of proxemic distance.

Performativity and vocal staging

Given that all forms of identity are socially constructed rather than a priori, I generalize a definition of performativity based on Thomas DeFrantz's notion of Black performativity, as "gestures of Black expressive culture, including music and dance, which perform actionable assertions" (DeFrantz, 2004, p. 67). My broader definition is thus that performativity consists of the repetition of performative actions that are denoted by a community as being appropriate for a particular aspect of identity (see also Butler, 1993, pp. 4–12). It is important to understand that performativity in pop music is enabled through technologies of music production and staging. Hawkins pointed out in analyzing Madonna that performativity and performance technologies are inextricably linked, "Behind her productions there is a technical gloss that highlights the striking traits of her aural and visual spectacle… This is rooted as much in musical style as performance design" (Hawkins, 2004, pp. 188–189).

Central to performativity in popular music is the voice, and many scholars have approached subjectivity, agency, and staging of voices in pop music. Although he does not implement hermeneutic approaches, Serge Lacasse categorizes in great detail the multitude of compositional and technical effects used in popular music staging (Lacasse, 2000). Moore's account of vocal staging connects the language of music technology and production with an interpretive methodology (Moore, 2012, pp. 101–118). Importantly, Moore emphasizes both the sonic characteristics of the voice and the lyrical text in his hermeneutic approach.[7] Going further, Freya Jarman uses a Foucauldian frame for understanding the role of music technology in identity construction, and distinguishes between "internal (physiological), external (recording, production), and power" technologies as frames for understanding the construction of voices, and by extension queer identities, in popular music (Jarman-Ivens, 2011, pp. 21–23). The congruences and juxtapositions of sound and lyrics (as well as instrument sounds) are critical to understanding the performance of identity.

Many have approached performativity by focusing on factors that extend beyond vocal staging. For example, Hawkins, in arguing that identity is as much part of a musicological as a sociological discourse, insists that "musical expression has a performative dimension from the outset" (Hawkins, 2002, p. 14). Similarly, Hansen unpacks performativity through hermeneutic analyses of audiovisual pop texts, and considers the various ways in which gender, race, ethnicity, sexuality, class, and other identifying elements are articulated in popular music and music video (Hansen, 2017). In a recent study, Danielsen and Hawkins illustrate that evidence for performative staging can be found when emphasizing the musical text as primary. This they demonstrate in Prince's personas and signature, which is shaped first and foremost through his virtuosity as a singer, guitarist, composer and producer (Danielsen & Hawkins, 2020). Looking beyond the musical text with a focus on racial subjectivity in Black rural feminist trap, Corey Miles cites Fred Moten to "situate Black performance in the Black radicalism tradition, suggesting it disrupts dominant discourses on Black subjectivity and is a form of resistance to

---

[7] This is, to an extent, in contrast to methods which focus solely on lyrics or solely on the sound of the voice. For example, Frith says that "the tone of the voice is more important… than the actual articulation of particular lyrics," and that this is "because it is the voice—not the lyrics—to which we immediately respond". (Frith 2004) (Frith, 2004)

objectification" (Miles, 2020, p. 47; Moten, 2003). What these authors remind us is that all aspects of identity are integral to the staging of pop performances.

I suggest that it is through the processes of vocal production and mixing that the artists' performativity is most impacted in the shift from stereophonic to immersive popular music. The recorded voice in popular music is recorded and processed through layers of reverb, compression, and other effects in ways that create a vocal sound which cannot exist in nature, but simultaneously fees acceptable and "natural" to listeners (Brøvig-Hanssen & Danielsen, 2013), and while this kind hyperreality is a common feature of popular music (Brøvig-Hanssen & Danielsen, 2016, p. 117; Lacasse, 2000, pp. 116–137), 3D music extends this hyperreality to an embodied interaction with the staged performer. 3D sound is of course a central feature of virtual reality, as can be experienced for example in Björk's 2019 release of the album *Vulnicura VR*.[8] However, I assert that the staging of the listener in 3D popular music does not require the high level of interactivity available in VR experiences but is an inherent aspect of the 3D sound format. In this way, the use of immersive music technologies in combination with existing pop vocal staging techniques has the potential to dramatically affect the possibilities for performers to stage their identities.

Subject positioning and designed perception

As much as artists perform identity through pop mixes, the real or assumed identities of listening subjects are also on display in musical texts. Here we turn to the concept of subject position, a term frequently used in media studies that describes the way in which media, through their formal properties, solicit particular responses by the interpreter. Quoting Clarke, "The notion of a subject-position is an attempt to steer a middle course between the unconstrained relativism of reader-response theory… and the determinism… of rigid structuralism" (Clarke, 2005, p. 93). My goal here is to understand the impact of immersion in 3D music on subject positioning. In a study I have undertaken with Hawkins, we have theorized how spatialities in immersive media necessarily stage listeners into the sonic environment in ways that can often be thought of as compositional: the listener of immersive audiovisual media is a staged object of

---

[8] Zachary Bresler and Stan Hawkins have written research about this very album, which is forthcoming.

the compositional design, their presence implying agential self-positioning.[9] This self-positioning is similar to that studied in video games. For example, Karen Collins has theorized deeply about interactivity in video game sound, and how sound in interactive media is the "method, material, and mediator of experience" (Collins, 2013, p. 13). Importantly, however, 3D music does not need the active interactive involvement of the listener to imply agency. Rather, my claim is that agency is impacted through the construction of the stage in and around the listener.

Illustrating subject positioning in compositional design, I turn momentarily to the verse of another track by The Weeknd, the 2016 hit 'Starboy'. Briefly, this song is a dark, braggadocio R&B track featuring the producers and performers Daft Punk. In the first verse, Tesfaye is heard singing close to his falsetto register but in his full voice and at a low volume into the microphone, using a technique that might be described as "cry" in Estill vocal technique (Steinhauer et al., 2017). Looking at the ways that other artists in pop and R&B such as Prince, Michael Jackson, Justin Timberlake, Pharrell Williams, and others have used this vocal styling, it is generally associated with an attitude of love or seduction.[10] However, juxtaposed against lyrics like "I'm tryna [sic] put you in the worst mood," and "Made your whole year in a week, too," it is clear that Tesfaye is instead engaging in a calm, sarcastic 'talking down' to the listener. Tesfaye's masculinity is on display in this verse, as well as a clearly straight-male subject position. In the last lines of the verse, he sings, "Main bitch outta your league too, ah / Side bitch outta your league too, ah." Here, Tesfaye seems to presume that the listener is male (and straight) and positions them as such by sizing up his masculine superiority through sexual conquest. As this example makes clear, any interpretation of the music requires a reading of the relationship between performer and listener, and various staged aspects of the identities of listeners, presumed or not, must impact on this perceived relationship.

Perhaps the most compelling argument about the relationship between 3D music technology and subject positioning is that immersive music more easily enables embodied interpretations. In her essay on embodiment in virtual reality, Morie suggests that embodied experience in immersive media (specifically virtual reality) necessitates an isochronic existence of the body "in both the real and virtual

---

[9] Forthcoming research by Bresler and Hawkins on the VR music videos released by Björk in 2019.
[10] An example of cry technique being used in a sexually/romantically intimate way can be heard in the 2010 song 'Hypnotize U' by N.E.R.D., which is sung by Pharrell Williams.

worlds" (Morie, 2007, pp. 127–128). Certainly, in all media there exists the possibility that subject positions allow the viewer to experience an alternative point of view, even to the exclusion of their own, a phenomenon described throughout cognitive science, for example, in the notion of the flow state described by Csikszentmihalyi (1990). Morie's notion of the bifurcated body that experiences both the real and virtual simultaneously is important because it informs us that immersive media experience, which certainly includes 3D popular music, comes with a unique type of subject positioning that involves the sensory experience of being surrounded with audio and/or visual environmental cues, which importantly exist simultaneously to those coming from the listening environment, as well as the metaphorical embodiment cues that exist in the text itself. As I have exemplified in the 'Starboy' example, the music of The Weeknd is replete with both implicit and explicit subject positions and performative stances. The use of immersive music technologies for dissemination such as Dolby Atmos, in these cases, serves to exemplify agency in these positions, granting the performer a more spacious environment in which to paint their identity and the listener new opportunities to engage and interact with the performance in ways that constitute new embodied experiences and subject positions.

Blinded by the lights: Analyzing The Weeknd in Dolby Atmos

On The Weeknd

Shrouding himself in mystery and anonymity early in his career on YouTube and SoundCloud, Abel Tesfaye was known in the beginning only by his stage name 'The Weeknd', and he has arguably contributed to transforming R&B and pop music since he emerged in 2011. Known for being a pioneer in the 'alternative R&B' style, The Weeknd is well on his way to becoming one of the most important pop icons of his generation. His music varies in style from polished number-one pop hits to dark, some might say utterly strange, R&B ballads. Like many of his peers, he makes constant reference to unashamed drug use, addiction, sexual encounters of every variety, and hesitation towards the whole affair of pop stardom. Listening to songs from throughout his career so far, such as 'High for This', 'The Hills', 'Can't Feel My Face', 'Starboy', and 'Heartless', a central theme runs through his work: ambivalence. Sometimes it is deeply coded, hidden behind the upbeat pop production of Max Martin or Daft Punk. Other times, it comes out quite explicitly as exemplified by this chorus lyric from his 2019 hit

'Heartless', "All this money and this pain got me heartless / Low life for a life 'cause I'm heartless."

The music of The Weekend is loaded with imagery and reference in ways that demand interpretation from listeners. In addition, The Weeknd has had much of his music remixed for Dolby Atmos 3D, making the simultaneous existence in stereo format useful for comparison. Here, I analyze the track 'Blinding Lights' by The Weeknd from the 2020 album *After Hours*. I approach the analysis explicitly in the terms of the model I presented earlier and work through each of the three conceptual frameworks in order.

Balance and proxemics

In the first verse of 'Blinding Lights' there are already points of comparison between the stereo and 3D versions in the spatial construction of the mix. The instrumentation in the verse is relatively sparse, consisting primarily of the lead vocal, kick and snare drums, and a simplistic synth bass. In the background is a heavily low-pass filtered saw-wave arpeggio, and some small percussion hits enter in the second half of the verse. To visualize the mix, I employ Moylan's "Perceived Performance Environment" diagrams, which are an apt way of modelling the perceived spatial layout of a mix in both stereophonic and surround sound music (Moylan, 2002). However, since the music we are analyzing is in 3D, rather than surround, and therefore contains sound in the height dimension, I have included a color-coding to the diagram to show relative elevation between sources.

Looking at PPE transcriptions of the verse, the stereo version is a clear diagonal mix structure, with the lead elements occupying the space directly in the center, while the filtered synthesizer is a spread image that occupies the majority of the stereo width while being perceived to be in the rear because of its relative volume and high amount of reverb. The lead vocal has a small amount of reverb and delay, and this is done in stereo and panned to the right and left of the image. In the Atmos mix, a standard front stage image can be heard in the verse, where the stereo structure is more-or-less recreated in front of the listener. However, there is some clear elevation panning, such that the bass and kick feel as if they are centered around the listener and somewhat lowered, while everything else excepting the lead vocal has been elevated. In effect, this creates room for those other voices to be more present in the mix without taking the space of the lead vocal, an effect that is particularly noticeable upon the entrance of the syncopated percussion sounds, which are noticeably louder than in the stereo version. Additionally, the vocal

delay and reverb envelopes the listening position from behind, again allowing for a much closer and immersive feeling without the need to reduce the amount of reverberation.
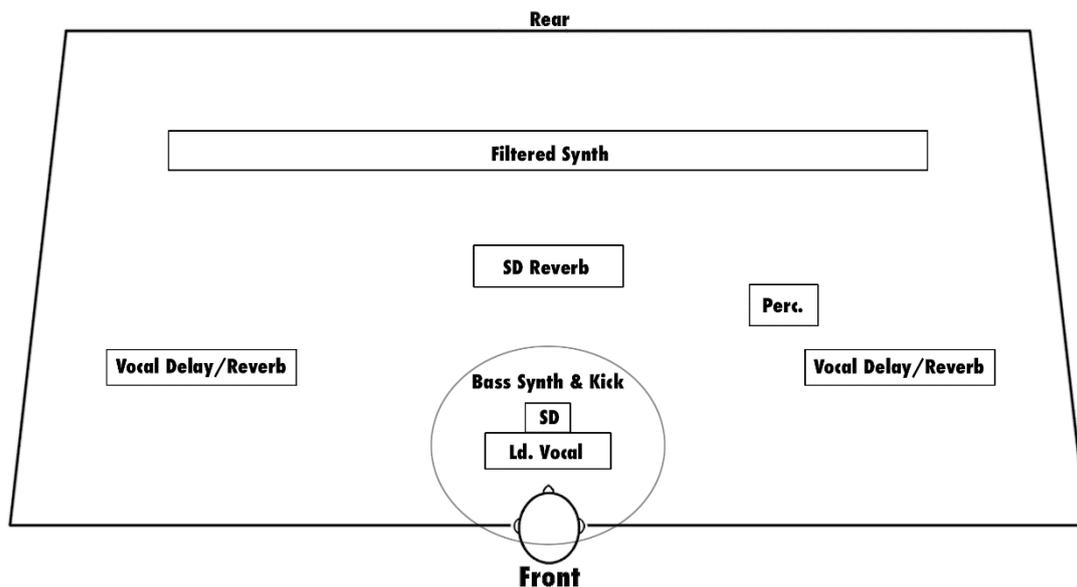

Figure 2: Verse of Blinding Lights, PPE transcription, stereo

Overall, the spatial difference between the two versions is that while the stereo version consists of a wide diagonal mix, the 3D version emphasizes the voice by surrounding the listener in it when possible while accentuating elevation in the background elements. This creates a wholly different set of proxemic relationships, as elements that are backgrounded in the stereo version become foregrounded in 3D. Additionally, the 3D format allows for different kinds of spatialization, such that nearly all elements can receive more reverb and delay, and that acoustic treatment can be panned in different directions to create the sensations of close-up sound sources that have long reverb times. While this is difficult to achieve in stereo mixing, it is quite common and much easier in 3D.
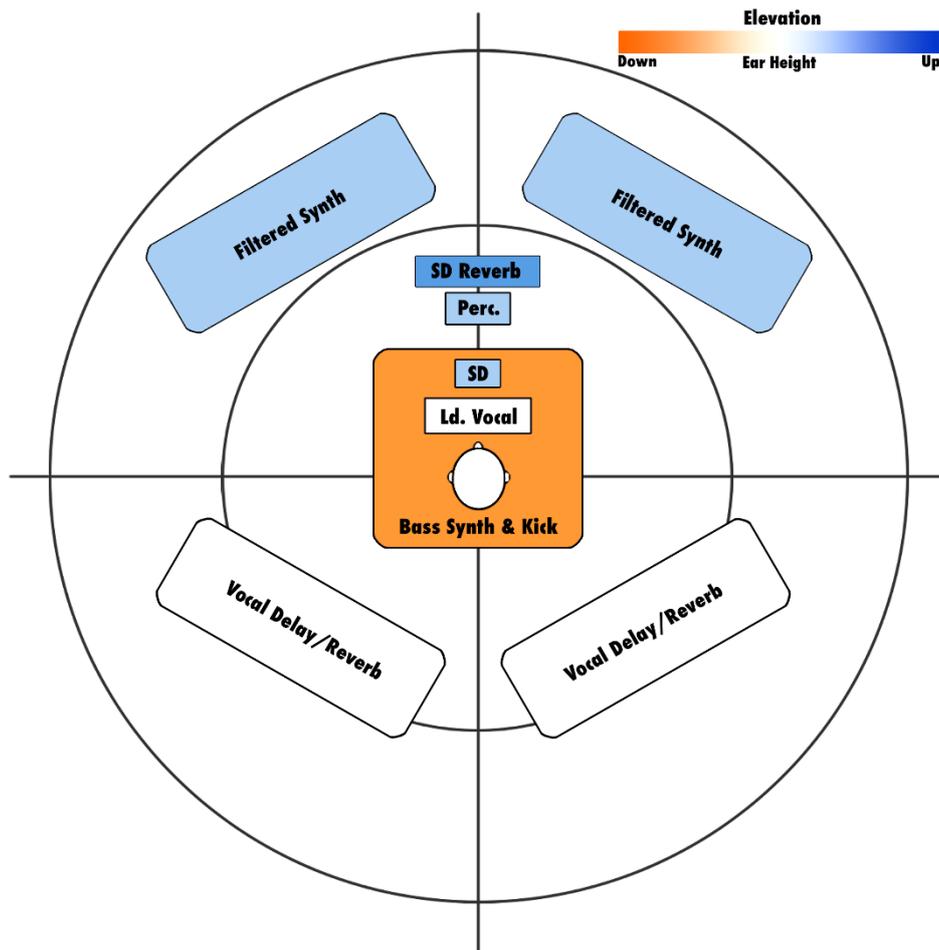
Figure 3: Verse of Blinding Lights, PPE transcription, Dolby Atmos

Performativity and vocal staging

From the first entrance of the voice in 3D, it is clear that the reverb and delay of the lead vocal are panned to the rear, while the lead voice itself has been left relatively acoustically dry and in front. Compared to the stereo, this has the effect of both increasing the spatial qualities of the voice while giving it even more size and immediacy with respect to the listening position. This approach of surrounding the listener in Tesfaye's voice is used throughout. Moving to the chorus, the voice is double (or triple) tracked, and the chorus of vocal lines and their associated reverbs and delays take up considerably more space in the Atmos mix, allowing the listener to hear more of the voice without obscuring the very impactful instrumental track. Cleverly, the instrumental hook of the song, played by a synthesizer reminiscent of the famous introduction to a-ha's 'Take On Me',[11] lands at the end of the chorus without any voice to compete against it. In both versions

---

[11] For a detailed analysis of this, see (Hawkins & Ålvik, 2018)

of the song, this synth is panned to the same place as the respective lead vocal, and in the Atmos version of the song the lead synth is given similar immersive treatment.

As with many contemporary pop productions, the voice in 'Blinding Lights' is heavily processed through various layers of compression, auto-tune, reverb, and delay, and in some parts is also double-tracked to add to the hyperreality of modern vocal music production. This comes through particularly in the verse, where it is clear that Tesfaye is singing at a low volume into the microphone, which in combination with the heavy compression allows us to hear mouth and throat characteristics in the voice which would otherwise be inaudible. Although he is singing in what sounds to be a comfortable range for his voice, the quality is thus both strained and intimate, suggestive of exhaustion and juxtaposed against the brightness and energy of the timbre and tempo of the track. This ambivalent sonic characteristic matches the exasperation of the lyrics:

> *I been tryna call*
> *I've been on my own for long enough*
> *Maybe you can show me how to love, maybe*
> *I'm goin' through withdrawals*
> *You don't even have to do too much*
> *You can turn me on with just a touch, baby*

The use of metaphorical language that obfuscates Tesfaye's relationships with women and with drugs is a play on words to which he frequently returns. Here, the use of the word "withdrawals" blurs our interpretation, as we are unsure if the one who will "show me how to love" is a lamentation of the loss of his lover or an admission that he doesn't feel himself unless high. Turning to The Weeknd's other music provides no answer, as he frequently refers to drugs as being like the women in his life, and vice versa.[12] This kind of vocal production in combination and juxtaposition with lyrics and instrumentals is one of the many ways in which Tesfaye stages his contradictory identity: he is tired, yet energetic; lonely, yet fulfilled; turned on, yet completely without desire.

---

[12] Probably the clearest rendition of this "drugs as women" metaphor comes in the song "Can't Feel My Face", in which Tesfaye apparently personifies his drug addiction as a toxic relationship which he cannot (or doesn't want to) end.

Later in the pre-chorus and chorus, we hear a double-tracked version of Tesfaye singing at full volume, occasionally with a cry-like quality when reaching the top of his register. Subtle as it is, Tesfaye is quite clearly using auto-tune throughout the song, cleverly leaning into the boundaries between notes at certain points in a way that effectively creates the illusion of a vocal break using the auto-tune processing. The voice in the verse also seems to be processed in parallel, so that the dry, compressed line is quite forward, while the reverb and delay version of the lead vocal are processed via a side-chain gate that 'opens up' the volume at the end of each line. In effect, this technique creates both the dry, forward voice while allowing the lush, layered reverb and delay to sit behind without bringing the lead back in space with it. In the 3D mix, this reverb and delay is panned mostly behind the listener, which again creates more space for the lead voice. In my reading, this draws attention to this effect in the rear even more than in stereo, reinforcing an interpretation that the singer is calling out into an empty void. Here, there is ample evidence that the 3D mix reinforces a level of ambivalence.

Turning to the short bridge (which comes in around 2:18), Tesfaye goes up in pitch near the limit of his range, singing noticeably louder and shifting towards a more public proxemic address:

> *I'm just comin' back to let you know (back to let you know)*
> *I could never say it on the phone (say it on the phone)*
> *Will never let you go, this time (ooh)*

Here, the repeated lyrics in parenthesis are echoes of the main line, and in the 3D mix, they are positioned behind the listening position, alternating between left and right. Additionally, they are filtered significantly such that they have a quality like that of a telephone or megaphone. In the stereo version, these are panned hard left and right and are also much louder than in the 3D mix. My perception in the Atmos mix is that the lead voice is 'closer', and the rear positioning of the 'background' echoes creates an almost 'devil-over-the-shoulder' feeling that positions the repetitions as internalizing thoughts to the subject position, rather than simply echoes or reiterations from the performer. I read this spatial structure as placing emphasis on the desperation in the lyrics while increasing the overall aesthetic feeling of being immersed in the voice—a very powerful moment in the song.
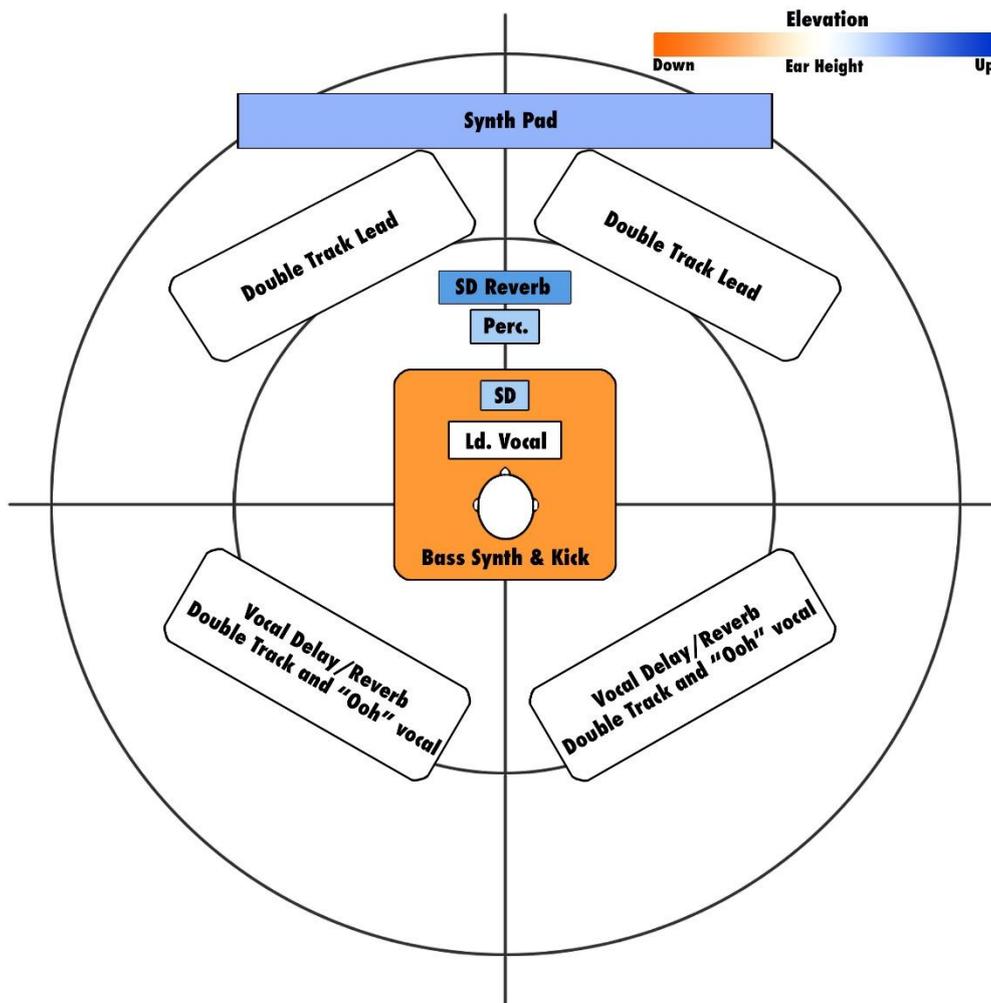
Figure 4: Chorus of Blinding Lights, PPE Transcription, Dolby Atmos

Implicit in The Weeknd's performativity is the matter of race, and while the music of 'Blinding Lights' offers few surface clues, turning to intertext, a picture emerges that allows us to gaze at this very important and interesting aspect of Tesfaye's staged identity. An immediate question in attempting to ascertain meaning in the lyrics is: what are the 'blinding lights'? The song itself is set in Las Vegas, the city that never sleeps, but he says, "Sin City's cold and empty / There's no one else around me," leading us to interpret the lights as being those of the late night/early morning strip. The outro of the previous track on *After Hours,* 'Faith'*,* has an almost seamless transition to 'Blinding Lights' and offers a different interpretative frame. Here we shift to an arrhythmic wash of atmospheric synths and a heavily filtered voice that sings slowly "I ended up in the back of a flashing car, with the city shining on my face. The lights are blinding me again." Many in the media have speculated this moment, and the lights in 'Blinding Lights' as a reference to a 2015 incident in which Tesfaye punched a police officer while being arrested in

Las Vegas.[13] In fact, in an interview in 2020 with the web magazine NME, Tesfaye clarified this point, speaking about the outro to 'Faith', he claimed that the period of his life in which this event occurred was "the darkest time of my entire life", and that the sirens in the background of the 'Faith' outro is "me, in the back of that cop car, that moment."[14]

Figure 5: The Weeknd in the introduction of the 'Blinding Lights' music video

At the beginning of the music video for 'Blinding Lights', a bloodied closeup of Tesfaye who is writhing in pain is visible, and in several live performances of the song for late-night television shows such as *Saturday Night Live* and *The Late Show*, Tesfaye performed with blood and bruise makeup and a large, white bandage across the bridge of his nose. He also performed in this makeup at the 2020 MTV Video Music Awards, where 'Blinding Lights' won awards for Best R&B Video and Overall Best Video, and he used his platform there to speak in solidarity with the Black Lives Matter movement, his bloodied face now clearly interpreted as a statement about the growing awareness of police brutality towards

---

[13] https://www.theguardian.com/music/2015/oct/23/the-weeknd-abel-tesfaye-avoids-jail-time-after-punching-police-officer, https://www.billboard.com/articles/columns/the-juice/6436581/the-weeknd-arrested-for-punching-las-vegas-police-officer, https://www.nme.com/news/music/the-weeknd-opens-up-about-2015-arrest-2643452, https://genius.com/The-weeknd-blinding-lights-lyrics.
[14] https://www.nme.com/news/music/the-weeknd-opens-up-about-2015-arrest-2643452

Black citizens in American communities. In fact, the context of the events of 2020, including the Covid-19 pandemic and the Black Lives Matter movement, offer many new interpretations of the song, from the lamentations of empty streets and lonely feelings to powerful illustrations of Black struggle.

From this it is clear that vocal staging is a primary tool that artists use to perform identity and persona, and this is made more vivid through the use of immersive music technology. By altering how the voice is positioned and processed to envelop the listener, the artist has reconfigured the stage to more include the listener and, at times, help the listener be immersed the vocal performance.

Subject positioning and perception

So far, I have delved into the ways in which The Weeknd has staged himself and the sonic structure that has been composed to support his identity. Now I want to turn briefly to the ways that the listener engages with immersive media, and how the subject position is constructed for them. At this point in the analysis, it is important to reiterate is that subject position is, by its very definition, ecological and highly dependent on which subject is being positioned. Clarke defines subject position as something that lies in the music: the "way in which characteristics of the musical material shape the general character of a listener's response or engagement" (Clarke, 2005, p. 92). However, as I have already problematized, the musical codes that are marked as relevant by one analyst may draw very different interpretive conclusions to those delineated by another. In other words, although here I intend to discover some generalities about subject positioning in this track, the analysis is unavoidably derived from my own hermeneutic self-positioning.

Unlike other music by The Weeknd, such as 'Starboy', subject positioning in 'Blinding Lights' is less defined and more broadly open to a wide variety of often contradictory interpretations. Sonically, the voice in the verses of 'Starboy' are dry—certainly compressed, but nearly without reverb and delay creating an extremely intimate 'spaceless' sound that reinforces the sensation that the singer is speaking directly to the listener. In 'Blinding Lights', the voice is treated with such a wash of reverb and delay that it feels as if the singer is in a huge and empty space, screaming into the void. While this is the case in the stereo version of the song, it is even more pronounced in the 3D mix since the voices are panned around the listener and the reverb tails are more present and easier to perceive even as they fade to the background. This effect is clearly discernible in the verse, as each line

is delivered over the course of about a measure, often followed by a measure or so of rest which is filled completely with reverb and delay that bleeds into the next line. In the chorus, the end word of each line is either "lights", "touch", "night", or "trust", and these hard, often sibilant consonant endings are perfectly timed with delay to create additional percussive movement on the offbeats.

In the lyrical analysis hitherto, it becomes clear that Tesfaye is purposefully obscuring the intended recipient of his words, forcing us to ask: to whom is he singing? Frequently he addresses a 'you', but I do not believe this is intended to address the listener as such. Rather, in this case the listener is an outside observer, and the subject of the singer's address is purposefully unclear, and, as I suggested earlier, lines like "I'm going through withdrawals" and "I can't sleep until I feel your touch" suggest a personification of addiction. One interpretation of the subject position is that the listener is transported to an observational stance that sees the performer who loudly laments to this nameless personification, and the heightened spatialization of the voice in 3D sound serves to further exemplify this position. Again, this is accomplished through the panning of the voice and its reverb all around the listener and the way in which that brings the main sound of the voice forward while creating a huge amount of empty-sounding space around the listener. A simultaneous interpretation is that the Atmos version allows for an embodied subject position—I hear the voice as dry because it is my voice, and I hear the echo all around me because it I am crying out to nobody. Regardless of the fact that these two subject positions are contradictory, it is nonetheless possible that they are not mutually exclusive, and one can hold on to them simultaneously. In fact, such contradictory spatial interpretations are part and parcel of popular music production. In terms of surreality in spatiality, Brøvig-Hanssen and Danielsen point out that "musical spatiality has a tendency to point the listener toward a real-world physical phenomenon even as it acts to undermine that reality" (Brøvig-Hanssen & Danielsen, 2016, p. 27). Likewise, the hyperreality of the immersive stage in 'Blinding Lights' creates a subject-position that is simultaneously embodied and distant, both extremely close and far away. In this way, the Atmos mix has reinforced the staging of ambivalence through the reconfiguration of the stage.

Conclusion

Increasingly, immersive and interactive editions of pop music are part of the mainstream media landscape, and as such it is important to put a focus on the ways

that such media impact on various interpretive aspect of pop texts. As I have attempted to demonstrate in this article, the relative differences between traditional stereo and immersive versions of pop songs lies not in the composition, but in the mix. While aesthetic features certainly change when moving between different forms of music media, structures that define the composition remain more-or-less consistent. In other words, any aesthetic changes are attributable primarily to the media format itself and can be seen as aesthetic features of the format. Analogizing to painting, aesthetic differences between the same image painted on different surfaces is a correlate of the aesthetics of the canvas, not necessarily the image.

In terms of proxemic distance perception, it seems that the amount of possible perceivable physical space has a large effect on perceptions of social distance. In 3D pop mixes, spatialization of musical elements and acoustic modelling serve to increase or decrease the apparent size and distance of performers and sounds, which in turn can create different possibilities for understanding and interpreting the musical content. Generally, acoustic modelling results in distancing (i.e. the intimate becomes the personal; the social becomes the public, and so on). Changes in proxemic perception will inevitably widen or narrow the possible meanings one can gleam from the text. Finally, frame of reference is critical in this context, as the perceived differences between stereo and immersive music will vary greatly based on the one's scope. Zoomed in on minute compositional detail, one sees little difference. However, zooming out towards meta-structures in space, interpretative stances, subject positions, and performativity, once can see that the reconfiguration of the performance stage creates new possibilities for all these aspects of music recording and performance.

Finally, as with all pop artists, The Weeknd carefully turns to staging to shape the perception and interpretation of various aspects of identity, including of course specific aspects such as race, ethnicity, gender, and class. The 3D music format, in this case Dolby Atmos, serves to reconfigure the stage. It changes the perceptions of relational space between the performer and audience, immerses the listener in the singer's identity through new approaches to vocal staging, and reinforces an interactive and embodied listener subject positioning. If music technology and its staging point to social and cultural self-positioning by artists and interpreters of popular music, then surely the dramatic ways that immersive and interactive media impinge on staging are important to consider; it is these kinds of media that continue to emerge more mainstream in the popular music sphere.

Bibliography

Auslander, P. (2008). *Liveness: Performance in a Mediatized Culture* (2nd ed.). Routledge.

Auslander, P. (2009). Musical Persona: The Physical Performance of Popular Music. In D. B. Scott (Ed.), *The Ashgate Research Companion to Popular Musicology* (pp. 303–315). Ashgate.

Brackett, D. (2000). *Interpreting Popular Music* (2nd ed.). University of California Press.

Brøvig-Hanssen, R. (2013). *Music in bits and bits of music: signatures of digital mediation in popular music recordings*. University of Oslo.

Brøvig-Hanssen, R., & Danielsen, A. (2013). The Naturalised and the Surreal: changes in the perception of popular music sound. *Organised Sound*, *18*(1), 71–80.

Brøvig-Hanssen, R., & Danielsen, A. (2016). *Digital Signatures: The Impact of Digitization on Popular Music Sound*. MIT Press.

Butler, J. (1993). *Bodies that Matter: On the Discursive Limits of "Sex."* Routledge.

Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. Oxford University Press.

Collins, K. (2013). *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. MIT Press.

Collins, K., & Dockwray, R. (2015). Sonic Proxemics and the Art of Persuasion: An Analytical Framework. *Leonardo Music Journal*, *25*, 53–56.

Csikszentmihalyi, M. (1990). *Flow. The Psychology of Optimal Experience. New York (HarperPerennial) 1990.*

Danielsen, A., & Hawkins, S. (2020). "The Right Amount of Odd": Vocal Compulsion, Structure, and Groove in Two Love Songs from Around the World in a Day. *Popular Music and Society*, 1–19. https://doi.org/10.1080/03007766.2020.1757814

DeFrantz, T. F. (2004). The Black Beat Made Visible: Hip Hop Dance and Body Power. In A. Lepecki (Ed.), *Of the Presence of the Body: Essays on Dance and Performance Theory* (pp. 64–81). Wesleyan University Press.

Frith, S. (2004). Towards an Aesthetic of Popular Music. In S. Frith (Ed.), *Popular Music: Critical Concepts in Media and Cultural Studies: Vol. IV* (pp. 32–46). Routledge.

Gerzon, M. A. (1992). Psychoacoustic decoders for multispeaker stereo and surround sound. *Audio Engineering Society Convention 93*.

Gibson, D. (1997). *The art of mixing: A visual guide to recording*. *236*.

Glasgal, R. (2001). Ambiophonics. Achieving physiological realism in music recording and reproduction. *Audio Engineering Society Convention 111*.

Hall, E. T. (1966). *The Hidden Dimension*. Doublday.

Hansen, K. A. (2017). Fashioning Pop Personae: Gender, Personal Narrativity, and Converging Media in 21st Century Pop Music. In *Department of*

Musicology: Vol. Ph.D. University of Oslo.

Hansen, K. A., Askerøi, E., & Jarman, F. (Eds.). (2021). *Popular Musicology and Identity: Essays in Honor of Stan Hawkins*. Routledge.

Hawkins, S. (2001). Musicological Quagmires in Popular Music: Seeds of Detailed Conflict. *Popular Music Online*.

Hawkins, S. (2002). Settling the pop score: Pop texts and identity politics. In *Popular and Folk Music*. Ashgate.

Hawkins, S. (2004). On performativity and production in Madonna's 'Music.' In S. Whitely, A. Bennett, & S. Hawkins (Eds.), *Music, Space and Place: Popular Music and Cultural Identity*. Ashgate.

Hawkins, S., & Ålvik, J. M. B. (2018). a-ha's "Take on Me": Melody, Vocal Compulsion, and Rotoscoping. In C. Scotto, K. M. Smith, & J. Brackett (Eds.), *The Routledge Companion to Popular Music Analysis: Expanding Approaches* (pp. 77–94). Routledge.

Jarman-Ivens, F. (2011). Queer Voices: Technologies, Vocalities, and the Musical Flaw. In P. T. Clough & D. R. Egan (Eds.), *Critical Studies in Gender, Sexuality, and Culture*. Palgrave Macmillan.

Lacasse, S. (2000). *"Listen to my voice": the evocative power of vocal staging in recorded rock music and other forms of vocal expression: Vol. PhD.* University of Liverpool.

Miles, C. (2020). Black Rural Feminist Trap: Stylized and Gendered Performativity in Trap Music. *Journal of Hip Hop Studies*, *7*(1), 44–70. https://doi.org/10.34718/kx7h-0515

Moore, A. F. (2001). *Rock: The Primary Text; Developing a Musicology of Rock* (2nd ed.). Ashgate.

Moore, A. F. (2012). *Song Means: Analysing and Interpreting Recorded Popular Song*. Ashgate Pub Co.

Morie, J. F. (2007). Performing in (virtual) spaces: Embodiment and being in virtual environments. *International Journal of Performance Arts and Digital Media*, *3*(2–3), 123–138. https://doi.org/10.1386/padm.3.2-3.123_1

Moten, F. (2003). *In the Break: The Aesthetics of the Black Radical Tradition*. University of Minnesota Press.

Moylan, W. (2002). *The Art of Recording: Understanding and Crafting the Mix*. Focal Press.

Moylan, W. (2012). Considering space in recorded music. In S. Frith & S. Zagorski-Thomas (Eds.), *The Art of Record Production: An Introductory Reader for a New Academic Field* (pp. 163–188). Ashgate.

Pink Floyd. (1973). *The Dark Side of the Moon*. Harvest Records.

Scott, D. B. (Ed.). (2009). *The Ashgate Research Companion to Popular Musicology*. Ashgate.

Senior, M. (2012). *Mixing Secrets*. Focal Press.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised Sound*, *2*(2), 107–126.

Steinhauer, K., Klimek, M. M., & Estill, J. (2017). *The Estill voice model : theory & translation*. Estill voice.

Discography

Pink Floyd, 'Money', *Dark Side Of The Moon*. Harvest. 1973
Salt-N-Pepa, 'Push It', Single. Next Plateau. 1987
The Weeknd, 'Starboy', *Starboy.* XO, Republic Records. 2016
The Weeknd, 'Blinding Lights', *After Hours.* XO, Republic Records. 2020
The Weeknd, 'Faith', *After Hours.* XO, Republic Records. 2020

Article 2 – "A Swarm of Sound": Audiovisual Immersion in Björk's VR Video *Family*

Zack Bresler and Stan Hawkins
Article is submitted and out for peer review as of submission

Introduction

Technological advances in pop music video productions have undergone significant changes in recent years, with performances increasingly spectacularized through the aid of new generations of camera devices and editing software. The advent of the internet has altered modes of consumption, sharing, and dissemination of the music video through the portals of Facebook, Spotify, Instagram, TikTok, Twitter, and YouTube (see Richardson, Gorbman & Vernallis 2013; Vernallis 2013; Korsgaard 2013, 2017; Burns and Hawkins 2019, p. 2). Music videos provide recourse for evaluating representations in new media technologies in a bid to understand the dynamic workings of artists from a range of disciplinary perspectives. For the purpose of this article, questions are raised that deal with aspects of immersion and interactive engagement. How do properties of compositional design in the virtual reality (VR)[15] music video function in establishing notions of space? What is the listener's role in multidimensional spatial environments? And how do the congruences of audiovisual sensory data enhance the causes and effects of sonic immersion?

Our starting point is to define audiovisual immersion as a pleasurable state of consciousness that is characterized by complete absorption, a result of the dialectic interactions between a viewing subject and compelling audiovisual experience. We have provided a model of the virtual audiovisual space (VAVS) that has as its objective to conceptualize experiences of audiovisual immersion in music. By extending ideas on 'virtual acoustic space' (Wishart 1996), this model emphasizes the relationship of visual imagery to sound and how this enhances agency, both on the part of the performer and viewer. Our goal is to consider sound and image

---

[15] Virtual reality (VR) describes a computer-generated, three-dimensional environment that can be experienced, explored and/or interacted with through the use of VR peripherals, such as isolating visual 3D headsets. Examples of VR systems include the Oculus Rift, HTC Vive, and Valve Index.

together as focal points for analysis in our understanding of virtual reality in music videos.[16]

In our study of the VR video, 'Family' (2019)[17] by the Icelandic pop icon Björk, we have considered the intermeshing audiovisual signifiers within a soundscape that enhances the sensations of immersion. Björk's progressive approach to technologies of immersion and interactivity has prompted similar scholarship, as evident in the studies of the "app album" *Biophilia* by Nicola Dibben (2013). Some of Dibben's claims about the app album can be applied to *Vulnicura VR*:

> *Biophilia* (re)introduced multimodality into digital audiovisual formats and used this to realize a creative vision of intuitive and embodied forms of music making and learning in which the natural word provides productive metaphors for emotional experiences and musical processes (2013, p. 699).

Working out the conditions of the multimodal virtual space in 'Family', we choose to concentrate primarily on the aesthetic effects of the VR performance. In the main, our model functions as a platform for considering the experiences of space and temporality within a highly active VR context; a context that functions as a staged environment that implies different things to every listener, culturally and socially. For instance, the design of sound and imagery might be perceived as surrealistic in one context yet entirely different in another. Hence, aesthetic experiences are contingent on a range of factors, and the analytical insights we provide are predicated upon personal interpretations and textual analyses.

We would suggest that the experience of audiovisual entertainment positions sound and image in the listener's memory. Lelio Camilleri's model of the sonic space (2010, p. 202) addresses this in terms of the 'localised space', the spectral space' and the morphological space', to which we add the 'aesthetic space', where sound and image synthesize in the audiovisual sensory experience. Such spatial

---

[16] Congruent with this goal, Anders Aktor Liljedahl has drawn attention to the way that studies of audiovisual media, including music video generally prioritise the visual and silence the music (2019, pp. 166-167).

[17] 'Family' was first released on the 2015 album *Vulnicura,* accompanied by 'moving album cover' featuring a short version of the song. The first VR video for 'Family' premiered in November 2016 at Harpa in Reykjavik (https://grapevine.is/icelandic-culture/art/2016/11/02/bjork-digital-opens-today/). The version discussed in this article is the video as it was re-formatted and re-mastered for consumer VR devices and released on the digital album *Vulnicura VR* in September 2019 on the Steam PC gaming platform.

combinations can be comprehended in terms of the aesthetics of sensory perception, creating the feeling of saturation. Integral to space in the music video are the properties of compositional design and the materiality of numerous 'stylistic and technical' codes (Hawkins 2002, pp. 9-12). Thus, audiovisual space accommodates the features and effects of sensory perception that instate the dramaturgy of a VR video performance. Given that music videos are contextualised within a mediascape, we also consider how *intermediality*, as defined by references, evocations, and techniques, impacts on VR productions.[18] One might posit that music videos are audiovisual compositional designs in themselves, their combined features mediated across any number of platforms during a performance. This would imply then that intermediality enables listeners to engage actively with the structural features of design.

In devising our VAVS model, we have been keen to highlight the attributes of "source bonding",[19] or the connection between heard sounds and their supposed causes (Smalley 1997, p. 110) that emanate also from shared experiences as they unfold through time and sensations of immersion. In the VR video we consider, it is as much the technological staging of space (in terms of texture, temporality, and gesture) as the musical features (rhythm, harmony, and melody) that define the 'aesthetic space'. Björk's VR performance comes across mysterious, if not scintillating since she breaks with many of the norms and traditions of the standard pop video format, which arguably becomes a metaphor for severing the constrictions of the conventional family unit. Our position is that 'Family' is derived from a pool of spatialities that denote a new audiovisual compositional domain and that enables the music to reach the viewer in a powerful and visceral way; that music immerses us within VR imagery is a highly personal affair.

The Virtual Audiovisual Space (VAVS)

If spatialities are integral to audiovisual contexts, being immersed in a music video is intermedial and multi-faceted. In Jem Kelly's words, the music video is "already a hybrid medium, comprising audio and visual forms and structures that intersect

---

[18] Intermediality as a term originated from intertextuality in 1983, which spawned a movement of intermediality studies led by German scholars, Aage Hansen-Löwe, Claus Clüver, Irina Rajewsky and Werner Wolf.

[19] Smalley's notion of 'bonding' goes beyond the idea of source-bonding used in this article. For Smalley, bonding simply is the way that sound and context are related, and source-bonding is one mode of this relation.

and interrelate in ways that can be described as intermedial" (2019, p. 220). We also adhere to the concept of 'multimodality', as theorised by Lori Burns, who considers "multimodality to comprise the artistic integration of multiple semiotic modes within one media text" (2018, p. 96). Part of what constitutes perceptions of agency is at the centre of the listening and viewing experience, and the video offers a glimpse of a specific context through a multimodal composite. The visualisation of the performance through immersion propels the viewer into a different interpretive space. Björk's VR video accomplishes this through the intermedial and multimodal relations of the sonic and the visual,20 which forms the basis of the audiovisual compositional design in the context of immersion in VR music video experience. Accordingly, we focus primarily on the context in which there is a "transgression of boundaries between what is conventionally perceived as distinct media" (Wolf 2015, p. 461).

Trevor Wishart's (1996) concept of virtual acoustic space (VAS) provides an in-depth insight into the compositional design of electroacoustic music and serves as an inspiration for our model of virtual audiovisual space (VAVS). In addition to Wishart's VAS concept, our study takes heed of Camilleri's sonic space (2010) and Denis Smalley's notion source-bonded spaces (2007, p. 38). In particular, we identify four dimensions in the audiovisual space: (1) sonic spaces:21 the environments in which sonic objects are placed, and their morphologies; (2) visual spaces: the virtual immersive imagery that constitutes what is visible (3) source-bonded spaces: the spaces in which the listening agent connects those objects to meanings through experience; and (4) aesthetic spaces: the abstract spaces where sound and image combine in the listener's memory to create meaning that transcends the source-bonding connections between the two (see Figure 1).

---

[20] We are acutely aware of the smudging of conceptual lines between intermediality and multimodality in this discussion so far, and therefore acknowledge the fact that all media comprise 'mixed media'. In our understanding, intermediality is the relationship between two media, such as music and imagery, and how they reference one another, while multimodality pertains to the application of variable literacies within one medium. For example, a music video performance involves the comprehension of language, culture, politics, and geography.

[21] Camilleri's model of 'sonic space' addresses the "space in which the [acousmatic] piece unfolds" (2010, p. 201). This three-dimensional model, which consists of localised space (the "space into which sounds are placed"), spectral space (the sensory understanding of timbre and disposition), and morphological space (the temporal aspect of space), accounts for the placement and disposition of sound objects as well as their morphologies, the ways in which such sounds are perceived temporally (ibid., p. 202).

In our model, the concept of sonic space pertains to the space created by auditory events and the way these events change in time and space. Indeed, the visual space denotes an extension of sonic space that accounts for the additional sense of vision. Just as sound can be described in terms of the positioning, disposition, and temporal unfolding of sonic objects, so can the visual space be understood in the same terms for visual objects. From this it is apparent that the sonic and visual spaces are not experienced or understood as distinct, although we argue it is productive to interpret them independently.[22]
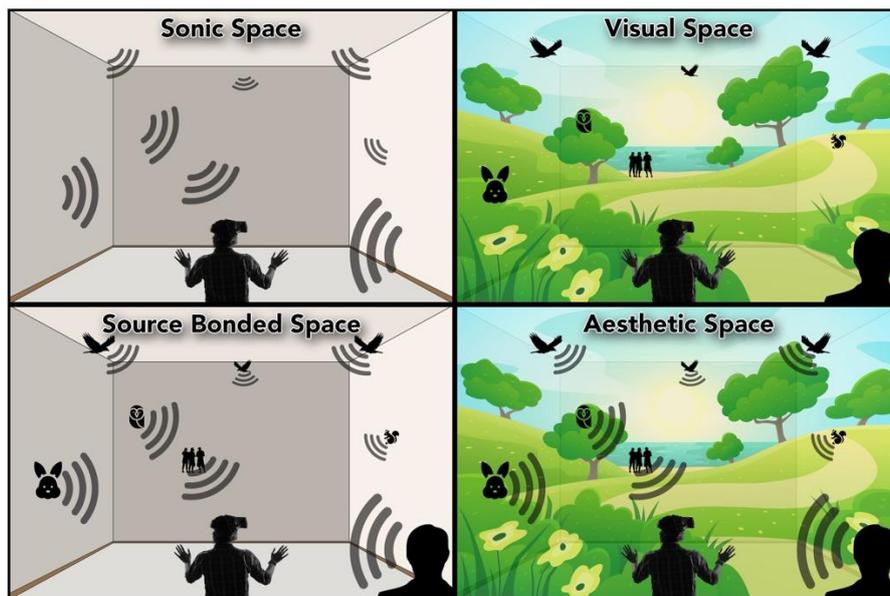

Figure 1: Virtual Audiovisual Space (VAVS)

Smalley has identified the spatialities created in the relationship between sound causes and their imagined sources as 'source-bonded spaces' (2007, p 38).[23] Accordingly, our model implements this to illuminate the ways in which the listener comprehends sonic and visual presentations via these assumptions.[24] As a

---

[22] For a comprehensive overview of the discourses around inter- and transmediality in music studies, see Werner Wolf's chapter "Literature and Music: Theory" from the 2015 De Gruyter *Handbook of Intermediality: Literature – Image – Sound – Music*. While our emphasis is more on intermediality, we have also considered that the music video might be framed within a *transmedial* context, where the media "unfolds across multiple media platforms, with each new text making a distinctive and valuable contribution to the whole" (Jenkins, 2006, pp. 95-96).

[23] Dwelling on Smalley's notion of 'source-bonding', as sounds are encountered they are processed in terms of their assumed causes, regardless of the fact that the formalization of the sound through recording, processing, and re-production through speakers has distorted the sonic image (1997, p. 110). Arguably, this has a corollary in Michel Chion's notion of 'causal listening' in the audiovisual experience (1994, pp. 25-28).

[24] As Smalley points out, this conceptualization of space is resonant with Lefebvre's notion of space as a social morphology. See Lefebvre's *The production of space,* originally published in 1974 and translated into English in 1991 by Donald Nicholson-Smith. Here, we relate this to acousmatic sound to suggest that

facet of immersive media, the 'source-bonded space' brings to the fore the listener's own interpretive perspective, and hence imports their agency into the model. In turn, this raises the idea of the artist's and listener's interaction during immersion, a significant aesthetic and experiential entity.

Given that the aesthetic space comprises a zone in which the multimodal experience is cognized and synthesized, it contains different modes of source-bonding that engage the viewer in ways that transcend either the sonic or the visual. In this sense, the aesthetic space is defined by the connections between sound and image that do not rely on exact sound-to-image source-bonding be understood. From this it is apparent that a higher-level of synthesis occurs simultaneously to the independent modes of sonic and visual understanding. Holly Rogers has suggested that the "audio basis, together with its continual motion, posits for the video image an existence in the musical sphere and vice versa… and its meaning no longer needs to be 'emergent' as it materializes, unified, at the moment of its creation" (2011, p. 410). Thus, in the context of immersive VR music video, aesthetic spaces signify a type of meta-spatiality, where new modes of sonic meaning arise (in contrast to the music without the video).

While source-bonded space relies on interpretation, it does so in the ecological rather than the hermeneutic sense, which, in our model, accommodates the domain of aesthetic space. This implies that interpretation in the source-bonded space does not necessarily require cognition, as this is done pre-consciously and corresponds to the viewer simply understanding the 'cause' of a sound and its relational context in the music. Eric Clarke has insisted that in this ecological mode of perception, "to hear a sound and recognize what it is… is to understand its perceptual meaning (2005, p. 7). Moreover, understanding the aesthetic space is a hermeneutic project. Accordingly, we adhere to Lawrence Kramer's call for "open interpretation", which "aims not to reproduce its premises but to produce something from them" (2011, p. 2). In this sense, source-bonded spaces represent the literal connections between sound and image, while the aesthetic space corresponds to the metaphors and meanings that we create in interpreting their connection.

Granted, none of these spatialities are mutually exclusive, implying that audiovisual cognition is only possible within the temporal framework of recalling

---

sounds produce the space they occupy through their understood relational attitude towards and with the listener.

the composition retrospectively as a singular event. As a result, notions of temporality are central to the audiovisual experience. Bearing this in mind, various strata are analysed that are deemed pertinent for inspecting the properties of immersion.

Immersion and Compositional Design

Returning to our concept of audiovisual immersion we now consider its inextricable ties to compositional design. Anders Aktor Liljedahl has stated that immersive and interactive music videos "suggest both an infinite set of outcomes and an enclosed range of possibilities," (2019, p. 183). To this we might add the immersive music experience, which creates endless possibilities of relating to compositional design. From one perspective, the 3D audiovisual experience 'de-idealizes' viewers as they become in themselves non-static and dynamic objects of the compositional design—a part of the music that experiences and creates meanings through interaction with the stage. Immersing oneself in surround sound and 3D music imagery entails grasping music production aesthetics more broadly. Just as the size and shape of the stage certainly matters, it is only a frame for accommodating the normative structures that define pop productions. As such, features of staging, when applied to and analysed in different formats, provide us with insights into the complexities of audiovisual immersion.

Quagmires of Immersion

Activating the term "immersion" necessitates closer inspection. Two critical points arise: first, "immersive audio" and "immersive media" have in recent years become buzzwords, largely used for marketing speakers, televisions, gaming systems, VR headsets, mobile phones, and any number of commercial electronics. In this sense, the term refers mainly to a type of media format.[25] In a broader audiovisual multimedia context, 'immersive media' primarily refers to virtual and augmented reality, accessible with devices such as the Oculus Rift or HTC Vive VR headsets. In studying VR and its immersive effects, Mel Slater has identified

---

[25] An example of an immersive audio format is Dolby Atmos, an object-based format that was developed for use in cinema and which has recently come into use for 3D music streaming on the HD tier of Amazon's Prime Music service. Atmos uses a standard surround sound configuration with an additional surround layer positioned a distance above the listening position. https://news.dolby.com/en-WW/182472-fall-in-love-with-music-all-over-again-with-dolby-atmos-on-echo-studio-and-amazon-music-hd

that immersive systems can be typologized based on their degree of immersive effect, characterizing virtual reality systems by their set of valid actions, that is, "the actions that a participant can take that can result in changes in perception or changes to the environment" (Slater 2009, p. 3550). Here, immersion is defined as "a property of the valid actions that are possible within the system" (ibid., p. 3551), and systems with more types and/or better qualities of valid actions are considered more immersive.

The second critical point is that, while Slater's notion of immersion is derived technologically, it is clear that immersion is also phenomenological. For example, it can be considered in terms of absorption, the state of consciousness that Graham Jamieson defined as "an effortless, non-volitional quality of deep involvement with the objects of consciousness" (2005, p. 120) and which is contrasted against an instrumental disposition which requires serious cognitive effort and planning.[26] Ruth Herbert goes further than Jamieson by applying the notion of absorption to the experiences of music listening, suggesting that "absorption and dissociation are best understood as processes that are subsumed within trance" (2011, p. 85). For the purposes of our definition of audiovisual immersion, we concur with Herbert's definition of "absorbed trancing", "characterized by imaginative involvement" that arises from "apparently passive yet still creative involvements such as listening to stories, listening to music, daydreaming, reading and imagining fiction, plus circumstances such as travelling on a train or being in a crowded place" (ibid., p. 134). Thus, retrieving an experience of being immersed need not be an audio or visual experience at all; it can be one of experiencing one's favourite music in any format, and it can just as well be another activity such as reading a book or taking a walk. The VR music video, however, combines many elements of absorption and trancing, enabling heightened sensory experiences that can lead to audiovisual immersion.

Immersion and Agency

Worth considering are the ways in which immersive media may more easily facilitate the immersive experience. One way that this occurs is through increased

---

[26] Immersion has been correlated to states of flow (Csikszentmihalyi 1990); immersive experiences are recalled in moments when slowing down allows for periods of intense focus. However, flow has been problematised in relation to immersion since it is "an extreme experience where goals, challenge, and skill converge. As such, flow is an all or nothing experience" (Sanders & Cairns 2010, p. 161).

agency on the part of the viewer. When watching a music video on a 2D surface, such as a laptop screen, the viewer assumes a passive role; engagement might well feel like interaction although the viewer is not staged in the same way as when entering a VR experience. Effectively, the 3D visual experience of VR engulfs the viewer. As with the Björk video 'Family', we have noted that this requires multiple viewings, since the decision to focus on one particular entity will inevitably lead to missing out on another. Notably, René Idrovo and Sandra Pauletto have extended ideas found in the work by Michel Chion (1994, pp. 90-91) and Rick Altman (1992, p. 60) on diegetic perspective in film sound, terming the "immersive point of audition" as "a sound design approach that aims to locate the audience on a specific point within the diegesis, and thus lures us to be transported into the story by providing an immersive representation of sound" (2019, p. 39). Extending this further, we would suggest that the agency of the viewer in a VR context places them *within* an audiovisual scene, signalling an *interactive* point of audition, whereby the viewer is not only placed on a point within the diegesis, but also has control over its perspectival transformations.

In the current research on user experience in media and games, concepts of immersion and engagement are critical to understanding the role of music. Engagement has been defined as our "ability to recognize a work's overturning or conjoining conflicting schemas from a perspective outside the text" and immersion as being "completely absorbed within the ebb and flow of schema" (Douglas & Hargadon 2000, p. 154). In video games, engagement through interactivity is a fundamental aspect of immersion and is typically considered in an embodied way, wherein "game controllers can become an extension of the body into the virtual world" (Collins 2013, p. 41). On the other hand, narratological approaches have often seen high degrees of agency as being antecedent to immersive experience, since it breaks the story into small, difficult to synthesize portions, while large complex stories require the rigidity of fixed, non-agential story structure (Douglas & Hargadon 2000, p. 155). We would argue that these seemingly contradictory notions of immersion are simply different classes of experience which constitute different modes of trancing. In general, audiovisual immersion in VR music videos is more like that in video games, where agency through interactivity is key, and where "the ability to move through virtual landscapes can be pleasurable in itself" (Murray 2016, p. 125).

In the *Vulnicura VR* music videos, as in many virtual reality experiences, one literally has the sense of being 'spaced out' through sheer immersion.[27] This results in temporary notions of de-virtuality, bridging the phenomenological gap between sensations of the real and the virtual. This relates to Jay David Bolter and Richard Grusin's notion that "virtual reality has become a cultural metaphor for the ideal of perfect mediation" (2000, p. 161). That is to say, that through the intensity of its means of mediation, it carries the potential to dissolve the very feeling of mediation. Complicating the boundaries between virtuality and reality is the notion of the digital, which promises that "our creative thoughts and imagination (i.e., the virtual) can be either transformed or *nearly* transformed into reality and actuality through digital means" (Rambarran 2021, p. 1, emphasis in original). As such, a major part of what constitutes this transformation in VR occurs through an interactive relation between taking "meaningful action and see[ing] the results of our decisions and choices (Murray 2016, p. 123). Again, the impact of agency on experiences of audiovisual immersion is critical.

In considering agency and immersion, we wish to stress the distinction between immersion and interactivity. While interactivity might be part of immersive media, it is not necessarily a part of immersive experience. Hitherto we have described audiovisual immersion as that sense of absorption within the media experience. Alternatively, interactivity is reserved for those instances in which the listener/viewer becomes an active creative agent. In addressing interactive installation art, Rogers states, "sound and image can be manipulated by visitors in order to create individual audiovisual pathways; or visitors in different location can be drawn together via technological intervention" (2014, p. 8). This suggests that a continuum of agency is possible within music and media, where at one extreme the listener is presented with a media at a distance, and at the other extreme they are transported into an interactive sound-world as a freely creative agent. Stereo music and 2D video are, in most contexts, closer to the former, while virtual reality is closer to the latter. However, as previously intimated, there are a number of other factors that contribute to the phenomenology of immersion, including the ability to engage meaningfully with the presented content in an extra-textual way. Accordingly, features of immersive media, such as surround and 3D

---

[27] Discourses on immersion and music listening are numerous. Some supplementary texts are worth mentioning here, including Tia DeNora's *Music in Everyday Life* (2000), Joel Krueger's article "Enacting Music Experience" (2009) and Simon Høffding's *A phenomenology of miusical abosorbtion* (2019).

sound and imagery, freedom of movement and position of the listener, and degrees of interactivity, create extra possibilities of immersive experiences, especially when content and context is made meaningful for the recipient.

Compositional design and perceptions of listening

Features of compositional design – a conceptual framework for describing how musical codes coalesce within a sound environment – lead to a holistic understanding of a track. Stylistic and technical codes can be utilized as part of a hermeneutic approach to music analysis (Hawkins 2002, p. 10-12), where the 'pop score' invariably comprises musical, social, and cultural objects that are coded and contextualized in such a way that the listener comprehends them as sonic representations of physical spaces and places. In addition, metaphorical, social, and cultural phenomena impact on our perceptions of compositional design and its structures.

Given that the central analytical framework for understanding the pop score is through its 'sound', then the use of the 3D sound stage in VR video has a significant impact. As the sound stage surrounds the listener, new subject positions, modes of performance, and proxemic relationships between performer and audience emerge (Bresler 2021). As we have highlighted in our analysis, this is often due to the simplicity of certain sounds appearing to emanate from unexpected locations, matching (or not) their visual counterparts in ways that push and pull the viewer's attention in multiple directions. In other cases, this is created by staging reverb, delay, and other secondary music processing behind the listener to create the feeling of particular acoustic spaces and places, or to literally surround the listener in a sea of voices or textures.

In applying notions of compositional design to the staging of audiovisual immersion, we are compelled to ask: where and how is the listener situated? In traditional media, the listener can often be perceived as metaphorically staged in an audience position with respect to the performance. However, this idea begins to disintegrate when creative and spatial formats 'surround' and engulf the listener. In immersive and interactive media, the stage is shared in an active way, with the listener positioned as a 'staged object' of the compositional design, their presence implying agential self-positioning. Certainly, the boundary between 'traditional' and 'immersive' media are not that distinct, and we do not intend to imply that films, stereo music, television, or any other form of media is incapable of creating such immersive experiences and staged subject positions. However, it is clear that

staging in VR is ontologically different from film with surround sound, for example, since the viewer expresses additional agency not only through their placement on the stage, but through their active participation in their own point of audition.

To demonstrate this we have undertaken an analysis of Björk's VR music video *Family*, released on the album *Vulnicura Virtual Reality* (2019).[28] Figure 2 provides a structural overview of the video and track through eleven discrete sections, with a focus on visual details, audio design, and the overall immersive effects (as we the authors encounter them). This table represents a semiotic analysis of the track, and functions as an aid to understanding the audiovisual processes and elements inherent in the VR music video.

Figure 2: 'Family' from Björk's Vulnicura VR, a detailed close reading (Next 2 pages)

---

[28] The researchers viewed the video on an Oculus Rift VR headset. At the time of writing, readers who are interested in viewing the video or the complete *Vulnicura VR* album will require this or a similar headset (such as a Valve Index or HTC Vive, and a PC computer with a suitable graphics processor. The album is, at this time, only available for purchase on the Steam PC gaming platform.

Screenshot / Time	Visual Details	Audio Design	Immersive Effects
(0:00 – 0:22)	The video commences in the dark, with a mysterious, purple, oblate luminous object. The viewer's digitized hands become visible as glimpses of flashing light reveal the environment as a tunnel-like cave, in which the viewer is moving with the object slowly forward.	The audio track begins with low strings uneasily blending between two notes a major 2nd apart. Very quickly we hear a loud, low-frequency impulse, followed by dissonant electronic sounds that resemble feedback and digital stutter. The impulses repeat regularly.	The first sensation is one of darkness, the appearance of a purple object, and sighting one's digital hands. The movement of exploration at the outset serves to establish the sound as immersive. It is immediately apparent to the viewer that both the visual and auditory aspects can be controlled through movement.
(0:23 – 1:36)	Björk, represented as a digitized body, flashes in and out—a translucent figure with child-like buns in her hair. We move with her slowly through the cave. As the flashing lights continue, tentacle-like structures appear behind the oblate object, out of which purple streamers begin to pour.	The sounds already established continue to along the same lines as Björk begins to sing. She starts with the line, "Is there a place, where I can pay respects, for the death of my family…" Continuing, the voice introduces a kind of call-and-response with Björk's voice replicated and panned to the rear of the soundstage.	At this point, the viewer will be aware that that their hands can be made to move in a kind of slow conducting pattern if the trigger is pressed on the controller. This also causes streamers to pour from the object to track the hands.
(1:37 – 2:16)	Björk, in sync with a lyrical cue, falls to her knees in front of the viewer.	The singer laments, "So where do I go, to make an offering? I fall on my knees." By now, the intensity and volume of the strings slowly increase in the background.	As Björk sings, vocal directionality is towards her physical manifestation. The contrapuntal lines of the voice, however, are panned away from the front, immersing the viewer completely. The voices to the rear seem at times distant as well, drawing the focus forward in the direction of movement.
(2:17 – 3:05)	At this point, the end of the cave comes into sight, as the music builds slowly towards a climax. The end of this section is as visible as it is audible. In the distance a sculpture apparition appears, albeit difficult to discern.	Breaking from the previous call-and-response structure, the two vocal lines merge together contrapuntally, as Björk sings "So where do I go, to make an offering, to mourn our miraculous triangle, Father, mother, child…" with each line, the number of harmony and counterpoint voices increases.	Entry of new vocal harmonies provide a sense of lateral imaging, as the lead vocal now assumes more space in the front. New vocal strands surround the listener, some of which are distant cries, others like whispers in the ear. Visually, tentacle structures now surround the viewer, constructing a kind of magical mothership that propels us through the cave.
(3:06 – 3:49)	The object in the form of a wound becomes more monotone and simplistic, yet still present alongside the translucent body figure. Surrounding the listener are several black and grey sculptures of Björk, bending backwards with her hands touching her feet, and rolling in that direction out of sync with one another.	At this point there is a dramatic change in the music in the form of a transition passage. The strings now play in an erratic, pizzicato Penderecki-esque style. The voice reaches a peak in a poignant outcry, "How will I sing us, out of this sorrow? Build a safe bridge, for the child, out of this Danger?"	All sense of directionality is temporarily lost as we are guided primarily by the changing directions of the lead vocal. Engulfed in the moving statues, lit only occasionally by strobe-like lighting bursts, this section is slightly disorienting. Especially on first viewing, the density of sound and visuals is overwhelming.
(3:50 – 4:49)	Suddenly, the visual field turns entirely white as one's eyes adjust to the intense daylight upon leaving the dark. Björk appears in front of the viewer, now larger than life and stylized in translucent pastel purple shades. Gradually, a magical, psychedelic Icelandic landscape is revealed: mountains in the background are offset by volcanic rock on the ground, and yellow northern lights across a purple sky.	As the erratic strings fade, they give way to a long, consonant tremolo on the high strings before transitioning to luscious steady chords. Right away, the high strings are supported by thick synth pads. Now the voice is notably calmer in tone, both musically and lyrically, as the material becomes lush, "I raise a monument of love. There is a swarm of sound."	As the climax is reached, the whiteness is at first blinding, and thereafter calming light with coloration is experienced. As Björk's body reappears and she sings, the viewer is reoriented towards her.

(4:50 – 5:14)	After a while, Björk's purple body disappears, and the landscape becomes more visible. Now the purples are replaced by dark greens and oranges. A black rocky sculpture like those seen earlier in the cave comprises the surface material, but with a dripping, purple "wound" vertically across the chest of the figure. Disembodied arms resembling those of the viewer's begin conducting movements above the sculpture.	The music repeats the previous phrase. The sound of wind starts to become audible, matching the heightened tangibility of the visual scene.	Musically, the strings and the synths which play in the same ranges enter from all directions, often seamless in their combination. The wind sounds also move past the listener, from one direction to another. These wind sounds are filtered in such a way that they lose their high end as it moves away. The effect is realistic.
(5:15 – 5:29)	Björk is 're-born' as she rises from the statue like the phoenix from the ashes. The object, now clearly a kind of wound on her chest, turns into a glowing source of light. The body becomes technicoloured with bright orange streamers now flow from the light in her chest.	Lyrical cue: "It will make us part of, this universe of solutions, this place of solutions, this location of solutions."	In this brief moment, the voice is solo, and experienced quite wide laterally. As Björk's body raises from the ground, it moves up continually. Once again, streamers pour out, with the viewer controlling their flow with the hands.
(5:30 – 6:09)	Having risen from the statue, Björk begins to walk slowly toward the viewer, as the viewer moves backward through the magical landscape. The medusa-like tentacle structure from the cave scene has returned, now framing the viewer from behind. As she walks, Björk is performing the same slow, conducting gesture with her hands.	Suddenly, the wind turns gusty as a cacophony of vocal harmony and counterpoint joins the lead voice. This point of multivocality becomes totally immersive as the visible singer walks slowly toward us.	Contrapuntal and harmonic vocal lines return, now totally consonant. They completely immerse the listener, the sensation being of warmth and comfort, rather than the fear and angst experienced in the cave. Visually, we are surrounded again by a purple tentacle-like structure, while the control over the streamers encourages dance-like motions.
(6:10 – 7:19)	At this stage, Björk's body has transformed its colour palette to various shades of deep red, purple, and orange. The colours resemble a sunset, the sky changing to reflect this event. She now looms larger than life, beginning to glide through the viewer through this section.	By now, the singing ceases, with high strings bending between pitches at a medium pace are clearly audible above the sound of the wind and the luscious synth pads, which sound as if they may actually be the filtered sound of synthesized voices.	In this section, Björk's body literally moves through the viewer, and in so doing prompts an urge to avoid this encounter. However, it is inevitable that she will subsume the viewer. For an instant, the viewer is encouraged to turn around and experience the scene from her perspective.
(7:20 – End)	As sight of the landscape dissipates, we are left with a deep purple and red haze. Björk's body, now hovering behind the viewer, slowly dissolves with the music, as the sets ends black.	As the piece draws to a close, the high strings slowly filter out, and are replaced by a sawtooth-style synthesizer playing the same line. The synthesizers and the wind gently fade out as the visuals fade into black.	Imagery and music dissolve into the sunset, at which point the viewer is rotated around Björk, who has walked through them. In the last moments she turns around to face the viewer before fading to darkness. Upon completion, when one removes the VR headset, they find themselves standing completely "backwards" of the starting position—a deliberate sense of disorientation seems to eb the objective here!

Immersion, Agency, and Aesthetics in Björk's VR video 'Family'

Björk belts out, "I raise a monument of love, there is a swarm of sound, around our heads, and we can hear it," as she reaches the climactic moment of the track 'Family'.[29] The lush shimmer of the Penderecki-like strings and the darting beats at this climactic point encapsulate the album's title, *Vulnicura*, namely, to be vulnerable and to be cured. This moment in the song comes across epic, a moment of transcendence when the mist lifts and the material gleams; the sonic landscape is ethereal, eloquently designed to create the sensation of a healing effect. Collaborating with Andrew Thomas Huang, Björk would experiment with digital VR technologies to produce a music video that vividly expressed her immersive experience. Huang has described how he designed the set and objects of the video:

> With a drawing and painting background, that's something I can do quite easily. It's really enriching, whereas shooting 360° video is more of a documentary-like workflow. For me, the 360° video is interesting because you are seeing the world captured as it is, untouched. Ideally with you erased.[30]

The effect of the motion-capture of Björk throughout the VR video is hyperreal, conjuring up notions of traversing magical landscapes (which were the actual landscapes used on the sets of 'Black Lake' and 'Stonemilker' shot in 360° video in Iceland). Baudrillard's theories of simulacra spring to mind when interpreting the engagement with representations of reality.[31] In a sense, hyperreality involves a simulation of reality and virtual immersion that operates more as real than the real (read: hyperreal); indisputably, Björk's VR performance creates this impression of a heightened reality. The narrative of traveling and searching is a veritable magical mystery tour, where the protagonist entices the viewer into her world by various means of identification. Compositional design, in both the imagery and music, elevates impressions of valleys, fjords, caves, open skies, and mountains. Designed by James Merry, the digital aesthetic of an artificial representation of nature is highly expressive. Björk's larger-than-life space-forms

---

[29] In chapter 10 of the Routledge research companion to popular music and gender (2017), Freya Jarman has undertaken one of the first studies into the phenomenon of belting out in popular music singing.
[30] https://www.vice.com/en_us/article/yp58dg/bjork-teases-family-virtual-reality-film-visuals
[31] See Baudrillard's essay 'The Precession of Simulacra', from *Simulacra and Simulation* (1981).

in 'Family' correspond to the viewer's own journey, where the sense of travelling in space enhances the digitalized spectacle of nature. We now turn to features linked to compositional design, in particular immersion and virtual reality, audiovisual creativity, impact of agency, and VR aesthetics These have direct correlation to the spatialities described within the VAVS model and form critical points of reference in our analysis.

Immersion and The Aesthetic Space

Music contributes to a powerful sense of presence in the VR experience, with the recording furnishing an aesthetic space. There is little doubt that Björk, as artist-composer, entered into this project with an acute awareness of this and a high degree of sonic spatiality. The video's narrative unfolds as part of a game-world where actions are played out by the main character and framed by psychedelic artwork. Immersion is achieved by a lateral sense of motion that is constantly fluid – the depiction of an Icelandic landscape as a dreamworld through which Björk travels blurs the distinction between fantasy and reality. At the climactic moment (3:50), the world transforms from a dark cave to a magical purple and yellow psychedelic-tinged surround, with Björk's digital body moulding into the same palate as her surroundings. If and when looking upwards, the viewer perceives what resembles a shimmer of yellow 'northern lights', spanning the purple background, signalling both a shift in the subjective perspective and the magical scenes of Iceland. Importantly, the temporal displacement of the sound image in the overall experience functions to generate various impressions of an active environment.

Immersion, in this instance, is mediated by Björk's agency and predicated upon a host of intricate details. In terms of the audiovisual space, Björk's gestures literally reach out to the viewer (for instance, from 5:30 onwards), beckoning them to make contact with her virtual hands. Useful here is William Moylan's notion of 'lateral imaging' that describes the placement of sonic objects in the sound stage, as well their perceived size and width (Moylan 2012, p. 176). In 'Family' Björk's larger-than-life presence at particular moments (e.g., at 3:06, 3:50, 5:30, 6:10) is contingent on expansions in the apparent size of her voice(s) in the 3D mix. The viewer is drawn into the aesthetic space through their own movement in the lateral plane; the result is that of feeling lost in space. This phenomenon is heightened prior to the 'cave climax' (3:06) when the absence of the singer in the forefront creates a confusing laterality causing the viewer to search their surroundings, and

also towards the end of the video (~6:10) as Björk's digitized body glides literally through the viewer. With subtle movement and interaction, the viewer increases not only their propensity for immersion through the agential space, but also their perception of the surrounding lateral spatialities created both by the artist and viewer.

From this we want to suggest that part of what constitutes aesthetic space lies in the perceived proxemic relationship between performer and viewer, since the viewer is in constant interpretive negotiation between their own subjectivity and that of the artist. There are specific ways in which the construction of the sonic space creates new proxemic relationships in the aesthetic space. For instance, the 3D mix in virtual reality allows for a placement of reverb and delay surrounding the listener, creating proxemics which are simultaneously perceived as intimate while retaining vast and lush reverb and delay profiles. This is certainly the case in the aforementioned 'cave climax' (3:06), where the reverb on Björk's voice is panned opposite to her position in the 3D mix. Later in the piece, there are moments when the counterpoint vocal lines are as loud or louder than the main vocal line which tracks the singer's digital body. These vocal parts are afforded various spatialisation profiles, from the feelings of whisper in the ear to distant repetitions of the lead voice's lyric and melody. In her analysis of the Björk song 'Vespertine', Dibben states that "the lyrics are simultaneously intimate and self-revealing such that they accomplish a striking alignment of the sensual with the spiritual" (2007, p. 176). Similarly, the lyrics in 'Family' suggest such tendencies, not least in the intimacy of Björk's vocal sound and its skilful panning in the 3D mix. As in much of her work, Björk turns to structures of intimacy through the minute details of her recorded voice, and in the 3D mix this is embodied in the background voices which become inner thoughts, whereby the cacophony of vocal textures has a sense of reflective and emotional inner dialogue.

Immersive sensations, such as those described above, establish a mutual space for performer and listener, the purpose being to communicate a sense of musical passage. Impressions of changing spatiality thus establish a holistic sensation of an environment that can pull us in any direction. In this case, agential space results from contrasting interfaces of colour, imagery, sound, bodily gestures, and above all, the three-dimensional surrounding of the viewer. In sum, the virtual reality experience of the narrative of 'Family' is conditional on the merging of physical

and cerebral interaction that works through a pop art aesthetic,[32] dependent on constant transformation and innovation.

Audiovisual Creativity: Voice and Visual Space

In the video Björk traverses a barren Icelandic landscape: imagery changes constantly align with musical events, the moment of transcendence occurring at 3:50, where the performer arrives at a 'gateway of enlightenment' suddenly drenched in a swarm of sounds. Structurally, the song's sections are relegated to visual happenings. Carol Vernallis (2019) has observed that song-sections unfold according to narrativity, and within the frames of each section of 'Family' Björk's agency can be assessed according to creativity. Overtly, she constructs her persona around a personal narrative that arguably possesses a high degree of authenticity.[33] On the function of the musical persona, Phil Auslander has stressed that an artist's appearance concerns "the visual dimensions of self-presentation, while manner has to do with the behavioural dimension" (2019, p. 96). With Björk, her persona is reinforced by a performance that is instantly identifiable as genre-specific in terms of trademark; her mannerisms and visual traits affirm an expression that is familiar to any of her fans, which directly mediates her very own digital signature. Ample opportunity to explore the imagination of Icelandic landscape, in tandem with her own quest for clarity, is on offer for the viewer.

Sonically, the treatment of the voice reinforces Björk's visually presented intimacy, both in relation to the viewer and to her connectedness with the imagery of Iceland on display. Although the motion-capture images of Björk's body are at times distorted, garishly coloured, and heavily stylised, the voice often retains a relatively dry and intimate quality. Close inspection in the sonic space, for example from around 6:00 until the end, discloses her voice dubbed and mixed in with luscious strings and synthesizer textures, coming across as heavily processed 'choirs'. In one sense, these choral textures serve as a connection point between the sonic and visual spaces, creating an aesthetic bridge from the sounds of strings to a surrealist landscape, all depicted by the voice. This also serves to confound

---

[32] By pop art aesthetic we are referring to both the surreal, profound and banal (Hawkins 1997) that stems from pop art's beginning in the mid to late 1950s where artists, such as Roy Lichtenstein, Andy Warhol, Jasper Jones, Tom Wesselman and others, derived their inspiration from subject matter found in everyday popular culture.
[33] For theories on personal narrative see Hawkins & Richardson 2007, Hawkins 2020.

the processes of source-bonding, since the sonic boundaries between choral overdubbing and string instruments are purposefully blurred.

Perhaps the most profound feature in the video is the sculpting of a space that is sonically expansive through technologies of spatialization. Tensions in the sonic materiality achieve different senses of space, where a wide array of sounds are constantly mobile; they are charged technologically through the details of production. Compared to the majority of her music videos, the VR technology employed in 'Family' arguably turns Björk into a 'virtual star', with a set of immersive qualities denoting a high degree of exceptionality. While the personal narrative might seem overt, there is an impression of a staged fictitious persona at work due to the 3D projections and communicative options open for the viewer to enter the set.

Impact of Agency: authorial intent

Common to Björk's *oeuvre* is a sense of full control of performance and production. In this sense, her 'authorial intent' within a transpersonal space (Hawkins 2002, pp. 15-16) can be assimilated against the authentic representations of her own attitude to performance. How then does the VR music video contribute to the relationship between viewer and singer? And what strategies are negotiated to facilitate a virtual sense of staging on the part of the listener as much as the performer? The impact of her performance results from the practice of signifying 'reality' in terms of the 360° video. Aspects of visual spatialization in 'Family' merge into a sonic soundworld where Björk's voice is foregrounded as intimate (Dibben 2012; Kraugerud 2020). Intimacy becomes a primary sonic device for drawing attention to the narrative and lyrical meaning. By creating a life-like persona in 'Family,' a sense of hyperembodiment[34] via the VR sensation is like being physically in touch with the artist. The multitude of positions offered up in the video are striking markers of agency, and one way to comprehend this is through multimodality where the composite of the performance is a result of different expressive modes. Burns has theorized this through 'expressive channels' or 'domains' that can be summed up as 'word-music-image' (2019, p. 184).

---

[34] Hyperembodiment is theorized by Stan Hawkins in an analysis of Rihanna's music video 'Umbrella' from 2007, where it is argued that an obsession with the look is conditional on technologies of musical production (See Hawkins 2013, p. 481). Also see Kai Arne Hansen's analysis of Beyoncé's sonic staging of the gendered body as a means to foregrounding hyperembodiment as a mechanism of digital fetishization (Hansen 2017).

Ultimately, Björk's embodied gestures guide the audiovisual aesthetics, brought into focus by the processes of production. As such, her corporeality is supported by the composite of word-music-image, which discloses an array of strategies.

One might say that the sense of journey in the VR experience entails a trajectory of author-induced imagery, inspired by the finely detailed audiovisual aesthetics. Björk's agency, a prime constituent and determinant of the compositional design, is aided by techniques of temporal regulation made realistic by close-up shots of her gestures and the merging of her with the viewer at specific points (for instance, at 6:10). Regulated integration with the viewer accomplishes a strong sense of identification, facilitating the pleasurable aspect of spectatorship. From 5:40 onward this is intensified as Björk begins a repetitive hand motion that resembles a kind of slow, ethereal conducting, or perhaps a sewing motion. This hand gesture can be performed and transferred to the viewer throughout the video.

Regulated by a button on the VR controller, in our case a pair of controllers for our Oculus Rift headset, the viewer's hands move in the same gesture and ensure interaction. Emphasis falls on Björk's agency as a performer is mediated through technologies of spatialisation as much as on the agency of the viewer. By the end of the recording, the impact of immersion is at its height as the viewer removes the headset; if they have 'followed' Björk's digitised character throughout, they discover they are facing the wrong direction—turned around 180 degrees from where they faced at the beginning. In such an instance of disengagement with the media, the viewer likely realises their own interaction with the song.

Worth emphasizing is the technology of viewing the VR video itself, which comes across with its own set of agential limitations. Jacquelyn Ford Morie has referred to the "bifurcated self", wherein "the act of emplacing one's body within the immersive environment signifies a shift into the dualistic existence in two simultaneous bodies" (2007, pp. 127-128). This is certainly the case in viewing the video on an Oculus headset and controlling one's digital arms with motion-sensing controllers—one is both the embodied character in a VR video and a person viewing the video, aware of the technological distance between the two selves, but feeling nonetheless connected to both.

VR Aesthetics

In probing further at VR aesthetics, we would suggest that the link between reality, hyperreality, and virtual reality is made tangible by optic arousal and strategies of representation. A wide palette of colours veer towards the florescent and garish at

times, enhanced by the use of lighting, which continually helps paint the environment, hence intensifying the staging of Björk's performance. Within an immersive VR environment, colours and lighting contribute to the perception of subjectivity. More specifically, nuances in the technical and stylistic coding of the compositional create visual impressions of light and shade. Correlating with timbre, texture and dynamics are finely regulated hues: blue merges with pink, orange turns into green, and so on. In addition, changes to sonic spatialization emphasize nuances of colour, signifying attitudinal and emotional content in the subject matter.

In assessing the full effect of the aesthetics of 'Family', we return to the question of the hyperreal and the digital simulation of reality both in the music and in the visuals. Narratively, there is a sense that Björk escapes the real world by journeying into the hyperreal one: the entire edifice of the digital production is reliant on her calling into question her own reality. It is within the confines of such spatiality that Björk's mission becomes most compelling. After all, the song is about her mourning the loss of her relationship to the US artist Matthew Barney in heart-wrenching lines, such as "Is there a place, where I can pay respects, for the death of my family." Ever so poignantly, her video stands as an unrivalled testament for expressing such pathos.

Conclusion

In concluding, we would like to return to the issue of staging and immersion in VR music videos and the cognizing of spaces, shapes, and designs. Dibben's in-depth study (2013) of Björk's approach to digitalization in *Biophilia* (2011) revealed the profound changes and effects a mobile app format had on the way people experience music. Introducing multimodality and interactivity to the experience of recorded music, the emergence of new aesthetic implications for the visualization and immersive modes of listening warrants attention. As Dibben insists, mobile music apps have formed a medium that offers interactive functions that lead to creative versions of Björk's songs. In opting for touch screens, Björk aimed for a new creative experience that combined technology, interactivity and nature. A distinctive feature of this was the integration of concept and aesthetic, which, first, encouraged a visualization of music in a way that encouraged "attentive listening to and playing with musical structures and processes," second, offered "a multimodal experience by virtue of touchscreen interactivity," and, third, presented "a curated experience of a coherent artistic vision" that was the result of

collaborative work (Dibben 2013, p. 688). A major consequence of the audiovisual relationships emerging from touchscreens was not only a renewal of modes of listening, but also a spontaneously embodied mode of engagement.

The powerful creative vision of collaborative music products by Björk would some years later be extended into the VR music format, which we have addressed in this article, where a host of new possibilities for comprehending compositional design are evidenced in our discussion of this format. As our model of virtual audiovisual space (VAVS) exemplifies, the listener's position and role merges with the act of staging a performance in an innovative format that extends the app album designed for mobile digital devices. One could argue that the audiovisual analysis of surround and 3D sound gets us to ponder over the developments and intricacies in music production on a broader scale. This is because the staging of sonic and visual devices in VR music videos allow for a greater sense of interaction between artist and fan, and in this dialogic space intertextual pathways are (re)invented and constructed. The conception here is that the new digital medium of music VR experience incorporates immersive functions that align pop music more to computer games. As with the user of games, the user of VR music videos has a wide scope to interact and perform along with the artist.

Ultimately, Björk's 'Family' VR video addresses features found in the artist's earlier work as we revisit her relationship between nature and technology through an acute practical engagement. Such fascinating technological innovations also give good cause for re-examining the normative structures that define pop dramaturgy, providing an opportunity to probe at the advances in technology and ponder over the future of new audiovisual aesthetics. Moreover, the multimodal aspect of 'Family' illustrates a coherent vision of compositional and performance design that has significant implications for understanding pop aesthetics and the phenomenon of music making. It is our hope that future studies will engage with the particularly nuanced phenomena of VR immersion, as a new generation of music video production continues to affect human development, agency, and creative expression.

References:

Altman, R. 1992. Sound Space. In: Altman, R. (ed.), Sound Theory / Sound Practice. London: Routledge.

Auslander, P. (2019). Framing Personae in Music Videos. In L. A. Burns & S. Hawkins (Eds.), The Bloomsbury Handbook of Popular Music Video Analysis (pp. 91-109). New York, NY: Bloomsbury Academic.

Baudrillard, J. (1994). Simulacra and simulation (S.F. Glaser, Trans.). Ann Arbor: University of Michigan Press. (Original work published 1981).

Bolter, J. D., & Grusin, R. (2000). Remediation: Understanding new media. Cambridge: MIT Press.

Bresler, Z. (2021). Immersed in Pop: 3D Music, Subject Positioning, and Compositional Design in The Weeknd's 'Blinding Lights' in Dolby Atmos. Journal of Popular Music Studies, 33(3).

Burns, L. A. (2018). Interpreting Transmedia and Multimodal Narratives: Steven Wilson's "The Raven That Refused to Sing". In C. Scotto, K. M. Smith, & J. Brackett (Eds.), The Routledge Companion to Popular Music Analysis: Expanding Approaches (pp. 95-113). New York and London: Routledge.

Burns, L. A. (2019). Dynamic Multimodality in Extreme Metal Performance Video: Dark Tranquillity's 'Uniformity', Directed by Patric Ullaeus. In L. A. Burns & S. Hawkins (Eds.), The Bloomsbury Handbook of Popular Music Video Analysis (pp. 183-200). New York, NY: Bloomsbury Academic.

Burns, L. A., & Hawkins, S. (2019). The Bloomsbury Handbook of Popular Music Video Analysis. New York, NY: Bloomsbury Publishing USA.

Camilleri, L. (2010). Shaping sounds, shaping spaces. Popular Music, 29(2), 199-211.

Chion, M. (1994). Audio-vision: Sound on Screen (C. Gorbman, Trans.). New York: Columbia University Press.

Clarke, E. F. (2005). Ways of listening: An ecological approach to the perception of musical meaning. New York: Oxford University Press.

Collins, K. (2013). Playing with Sound: A Theory of Interacting with Sound and Music in Video Games. Cambridge: MIT Press.

Csikszentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience. New York, NY: Harper Perennial.

DeNora, T. (2000). Music in everyday life. Cambridge: Cambridge University Press.

Dibben, N. (2007). Subjectivity and the Construction of Emotion in the Music of Björk. Music Analysis, 25(1), 171-197.

Dibben, N. (2012). The Intimate Singing Voice: Auditory Spatial Perception and Emotion in Pop Recordings. In D. Zakharine & N. Meise (Eds.), Electrified Voices Medial, Socio-Historical and Cultural Aspects of Voice Transfer (pp., 107-122). Göttingen, DE: V&R unipress GmbH.

Dibben, N. (2013). Visualizing the App Album with Björk's Biophilia. In C. Vernallis, J. Richardson, & A. Herzog (Eds.), The Oxford Handbook of Sound and Image in Digital Media (pp. 682-704). New York: Oxford University Press.

Douglas, Y., & Hargadon, A. (2000). The pleasure principle: immersion,

engagement, flow. Proceedings of the 11th ACM on Hypertext and Hypermedia. San Antonio, TX.

Hansen, K. A. (2017). Fashioning Pop Personae: Gender, Personal Narrativity, and Converging Media in 21st Century Pop Music. (Ph.D). University of Oslo, Norway.

Hawkins, S. (1997). The Pet Shop Boys: Musicology, masculinity and banality. In S. Whiteley (Ed.), Sexing the Groove. London: Routledge.

Hawkins, S. (2002). Settling the Pop Score: Pop texts and identity politics. Burlington, VT: Ashgate.

Hawkins, S. (2013). Aesthetics and Hyperembodiment in Pop Videos: Rihanna's 'Umbrella'. In J. Richardson, C. Gorbman & C. Vernalis (Eds.), The Oxford Handbook of New Audiovisual Aesthetics (pp. 466-482). Oxford: Oxford University Press.

Hawkins, S. (2020). Personas in rock: "We Will, We Will Rock You." In A. Moore and P. Carr (Eds.), The Bloomsbury Handbook of Rock. New York, NY: Bloomsbury Publishing USA (forthcoming).

Hawkins, S. & Richardson, J. (2007) Remodeling Britney Spears: Matters of Intoxication and Mediation. Popular Music and Society, 30(5), 605–629.

Herbert, R. (2011). Everyday Music Listening: Absorption, Dissociation and Trancing. Surrey, UK: Ashgate.

Høffding, S. (2019). A phenomenology of musical absorption. London: Springer.

Idrovo, R. & Pauletto, S. 2019. Immersive Point-of-Audition: Alfonso Cuarón's Three-Dimensional Sound Design Approach. Music, Sound, and the Moving Image, 13, 31-58.

Jamieson, G. A. (2005). The Modified Tellegen Absorption Scale: A Clearer Window on the Structure and Meaning of Absorption. Australian Journal of Clinical and Experimental Hypnosis, 33(2), 119-139.

Jarman, F. (2017). High Notes, High Drama: Musical climaxes and gender politics in tenor heroes and Broadway women. In S. Hawkins (Ed.), The Routledge Research Companion to Popular Music and Gender (pp. 137-151. New York: Routledge.

Jenkins, H. (2006). Convergence culture: where old and new media collide. New York: New York University Press.

Kelly, J. (2019). The Palimpsestic Pop Music Video. In L. A. Burns & S. Hawkins (Eds.), The Bloomsbury Handbook of Popular Music Video Analysis (pp. 219-233). New York, NY: Bloomsbury Academic.

Korsgaard, M. B. (2013). Music Video Transformed. In J. Richardson, C. Gorbman & C. Vernalis (Eds.), The Oxford Handbook of New Audiovisual Aesthetics (pp. 501-521). Oxford: Oxford University Press.

Korsgaard, M. B. (2017). Music Video After MTV: Audiovisual Studies, New Media, and Popular Music. New York and London: Routledge.

Kramer, L. (2011). Interpreting Music. Berkeley: University of California Press.

Kraugerud, E. (2020). Come Closer: Acousmatic Intimacy in Popular Music Sound (PhD thesis, University of Oslo).

Krueger, J. (2009). Enacting musical experience. Journal of Consciousness Studies, 16(2-3), 98-123.

Liljedahl, A. A. (2019). Musical Pathfinding; or How to Listen to Interactive Music Video. Music, Sound, and the Moving Image, 13(2), 165–85.

Morie, J. F. (2007). Performing in (virtual) spaces: Embodiment and being in virtual environments. International Journal of Performance Arts and Digital Media, 3, 123-138. doi:10.1386/padm.3.2-3.123_1

Moylan, W. (2012). Considering space in recorded music. In S. Frith & S. Zagorski-Thomas (Eds.), The Art of Record Production: An Introductory Reader for a New Academic Field (pp. 163-188). Surrey: Ashgate.

Murray, J. H. (2016). Hamlet on the Holodeck: The Future of Narrative in Cyberspace (2 ed.). New York: The Free Press.

Rambarran, S. (2021). Virtual Music: Sound, Music, and Image in the Digital Era. New York: Bloomsbury Academic.

Richardson, J., Gorbman, C., & Vernallis, C. (Eds.). (2013). The Oxford Handbook of New Audiovisual Aesthetics. Oxford University Press.

Rogers, H. (2011). The Unification of the Senses: Intermediality in Video Art-Music. Journal of the Royal Musical Association, 136(2), 399-428.

Rogers, H. (2014). Spatial Reconfiguration in Interactive Video Art. In K. Collins, B. Kapralos, & H. Tessler (Eds.), The Oxford Handbook of Interactive Audio (Online) (1 ed.). Oxford: Oxford University Press.

Sanders, T., & Cairns, P. (2010). Time perception, immersion and music in videogames. Paper presented at the Proceedings of the HCI International 2009, San Diego, CA.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of The Royal Society B, 364, 3549-3557.

Smalley, D. (2007). Space-form and the acousmatic image. Organised Sound, 12(1), 35-58.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. Organised Sound, 2(2), 107-126.

Vernallis, C. (2019). Writing about music video. In L. Patti, ed. Writing About Screen Media, New York and London: Routledge.

Vernallis, C. (2013). Unruly Media: YouTube, Music Video, and the New Digital Cinema. Oxford: Oxford University Press.

Wishart, T. (1996). On sonic art (2nd ed.). (First edition 1985). Amsterdam: Routledge.

Wolf, W. (2015). Literature and Music: Theory. In G. Rippl (Ed.), Handbook of intermediality: Literature–image–sound–music (Vol. 1). Berlin: Walter de Gruyter GmbH & Co KG.

Audiovisual Reference:

Björk. (2019). Vulnicura VR [VR Album]. UK: One Little Indian, Analog Studios. Available on Steam: https://store.steampowered.com/app/1095710/Bjrk_Vulnicura_Virtual__Reality_Album/ (Downloaded Oct. 2019).

Article 3 – Pop Music Diegesis and the 360º Video

Zack Bresler

Article is submitted and out for peer review at time of submission

Introduction

In this essay, I build on existing studies into music video and immersive media[1] by asking how immersive pop music video productions can shape the narratives that audiovisual pop texts attempt to illustrate, which I suggest works through technologically enabled agency and immersion. Ultimately, this work uses an interdisciplinary framework to suggest that so-called immersive media, in this case 360º pop music videos, situate the viewer on various levels within the narrative structure of music video, thus allowing for different modes of narratology and meaning in the agential space. Moreover, I want to ask: what are the audiovisual features that enable immersive experience in immersive media, and how do these forms of immersive media elicit subject positions differently from traditional films, recorded tracks, and music videos?

Creators of pop music productions often operate within narrative structures, conveying ideas through audiovisual storytelling. Part of the unfolding of a music video occurs in the "aesthetic space", where sound and image synthesize hermeneutic positions that are unique to their confluence (Bresler & Hawkins, 2021). In addition to source-bonding, the phenomena whereby sounds are associated with their supposed causes as they either appear on-screen or in the memory of the listener (Smalley, 2007, p. 38), the aesthetic space is formed in the viewer's interpretation, within which sound and image are connected to abstract feelings, intertextual sources, and deep personal meanings. To this, I add the agential space, explicitly suggesting that it is through interactivity that the viewer is granted a role in the diegesis of a music video.

For the present study, I choose to focus on 360º music videos, which are a form of virtual reality videos that are available to stream via platforms such as YouTube and Facebook. In short, these videos are captured or digitally constructed using a cylindrical video frame that is navigated in either in a head-mounted VR display

---

[1] See Bresler (2021); Bresler and Hawkins (2021); Burns and Hawkins (2019); Burns and Woods (2019); Dibben (2013); Hansen (2019); Jirsa and Korsgaard (2019); Kelly (2019); Korsgaard (2019a); Liljedahl (2019); Morie (2007); Rambarran (2021); Ryan (2001); Vernallis, Herzog, and Richardson (2013); Walther-Hansen (2015); Winters (2010).

(such as an Oculus), a mobile phone in augmented reality mode or in a mobile phone headset (such as a Google Cardboard), or simply through clicking to navigate on a mobile phone or computer screen. This format is chosen since 3D and 360º media offer an easily demonstrable case for the viewer's role in the diegesis of music video. I contend, however, that the findings are applicable to traditional music videos and even acousmatic music recordings. Furthermore, I advocate for the consideration of diegesis and narratology in popular music and music video analysis generally.

Developing on existing research within the field of popular musicology, I propose a two-fold hermeneutic framework which I call *pop music diegesis*, which relies on two aspects of engagement with interactive media: agency and immersion. I argue that a particular mode of each of these concepts is operational in the majority of 360º pop music videos, which I term as *navigational agency* and *diegetic immersion* (figure 1).
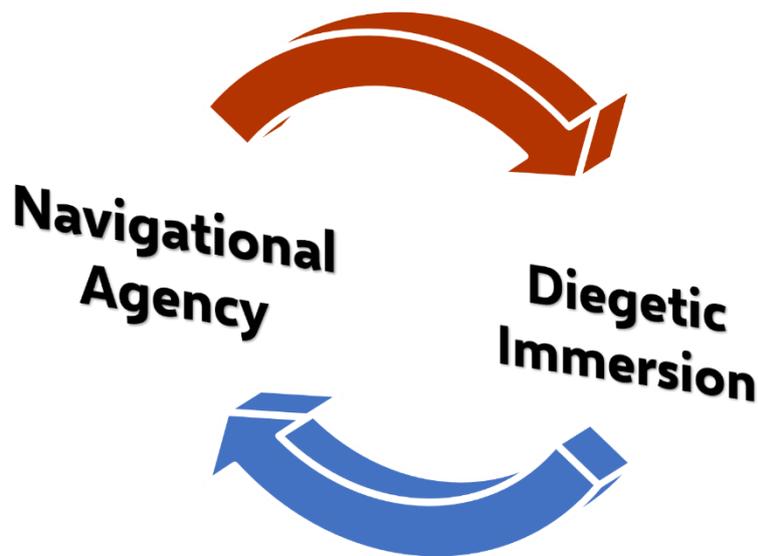

Figure 1: Pop Music Diegesis in 360º Videos

Navigational agency refers to the degree to which the viewer has control over their movements, while diegetic immersion refers to the degree to which the viewer has (or doesn't have) a defined and participatory role within the narrative structure. This framework is demonstrated through the inclusion of various examples from

four 360º pop music videos[2]: "Life Support" (2018) by Taryn Southern, "Revolt" (2016) by Muse, "The Hills remix" (2015) by The Weeknd feat. Eminem, and "Stor Eiglass" (2015) by Squarepusher. These videos are all freely available on YouTube and combine to represent a wide array of production techniques and narrative structures which demonstrate the various degrees of pop music diegesis. Aptly, they demonstrate the efficacy of interpreting music video through this framework.

Pop Music Diegesis

In media and film scholarship, it has been common to reference sound and music with respect to a film's diegesis, that is, the internal, logical space of the film's story world. The terminology is often cited to Claudia Gorbman and is rooted in narratological literature studies. In Gorbman's application, diegetic sound is that which emanates from the story world itself (i.e. dialog, sound effects), while nondiegetic sound is supporting the narrative for the viewer but is not "heard" as such from "within the scene" (i.e. soundtrack music) (Gorbman, 1980, p. 197). This dichotomy between diegetic and non-diegetic has been problematized by most film scholars since. For example, Ben Winters has argued for the essentiality of much non-diegetic music and sound to the "identity of the fictional narrative space presented in film" (2010, p. 230). Winters ultimately argues for the usefulness of the terminology, suggesting the term "intradiegetic" as the broad category for sounds which are fundamental to narrative structure—and thus central to the film's diegetic frame—but which are not implied to exist in the fictional story world as such (2010, pp. 237-238).

At this juncture it is critical to make note of the theoretical and disciplinary challenges of discussing diegesis at all in relation to music video. In film studies, narrative is often pitted against spectacle as its opposite. For example, Andrew Darley claims:

> …in critical studies of the dominant cinema institution, centred upon analysis of classical narrative films, attention has most frequently focused on the 'tension'

---

[2] "Life Support" by Taryn Southern: https://www.youtube.com/watch?v=LWl9Oi2NHps;
"Revolt" by MUSE: https://www.youtube.com/watch?v=91fQTXrSRZE;
"The Hills remix" by The Weeknd feat. Eminem: https://www.youtube.com/watch?v=2fhjdtQDcOo;
"Stor Eiglass" by Squarepusher: https://www.youtube.com/watch?v=6Olt-ZtV_CE

between the narrative dimension and the visual dimension, that is, between identifying with characters, being absorbed in a fictional world and following the plot on the one hand, and the pleasures involved in looking at images on the other (Darley, 2000, p. 104).

Darley continues that "spectacle is, in many respects, the antithesis of narrative… [it] halts motivated movement" (ibid.). This is concerning for this study, in particular since the aesthetics of music videos are often described *primarily* in terms of spectacle (Ålvik, 2017; Auslander, 2008, 2021; Burns, 2016; Hawkins, 2002, 2009, 2016; Korsgaard, 2013, 2019b; LaFrance, 2013). However, music videos defy categorization. Korsgaard has insisted that "rather than comprising a unified field, music video is actually defined by its very heterogeneity, its wide range of different audiovisual expressions" (2017, p. 37). In terms of narrativity, Vernallis has reminded that "music video presents a range all the way from extremely abstract videos emphasizing color and movement to those that convey a story" (2004, p. 3). Thus, one might surmise that there are as many genres of music videos as there are of music. I have not the time or space here to flesh out a typology, however my argument is that while music videos are often spectacular, many have varying degrees of traditional narrative structure that make diegetic perspectives relevant.

I argue that any perceived conflict between narrative and spectacle is ultimately semantic, rooted in definitions of narrativity and diegesis that include primarily classical narrative elements like characters and plots. In film studies, a similar debate has taken place regarding the role of special effects in films, in particular the grand digitally constructed visual spectacles in movies like Jurassic Park, Avatar, and Titanic. Aylish Wood has pointed out that such visual effects, such as the detailed digital reconstruction of the Titanic, "operate at another dimension of the narrative… that places a particular emphasis on the story of the fall of this technological giant" (2002, p. 372), and that it is the overlooking of this other dimension that "leads commentators to argue that spectacle interrupts narrative" (ibid.). Similarly, music videos make use of the "musicalization of vision", whereby images are "shaped according to and respond to different musical parameters" (Korsgaard, 2017, p. 65).

My argument is that while the images of most music videos are undoubtedly spectacular in the sense that they exist primarily for the pleasure of viewing them, this does not mean they cannot be diegetic, similarly to the way digital effects in

spectacular films create immersive narrative points of entry. Said differently, the spectacular elements in digital special effects and in the musical presentation of image in music video can be seen as diegetic because they operate on the level of world-*building*, as opposed to world-explaining or world-developing. Moreover, abstract as they may be, stories with low levels of classical narrativity are still stories, and the open-ended audiovisual design of music videos encourage viewers to read a multitude of narratives that explain the meanings of pop songs.

Considering diegesis from within the discipline of critical musicology, Walther-Hansen has theorized on the "phonographic diegesis" of pop music recordings, which is a kind of typology of recorded music staging centered around the idea of diegetic, meta-diegetic, and extra-diegetic sounds (2015). He approaches this by focusing on the edge-cases, wherein the diegetic framing changes through the course of the track, thus exposing diegetic boundaries. While Walther-Hansen is concerned primarily with sound recordings, I suggest that there is a further distinction to be made when considering the diegesis of a music video, and furthermore an immersive music video, as new narrative interpretations will surface in the aesthetic space as the music is made audiovisual, and even more through the agential space as the viewer is stage within the video itself. Importantly, while considering the diegetic framing of 360º music videos in particular may seem a niche project, like the examples in Walther-Hansen's study, it serves as an easily accessible edge-case for understanding the motivations, technologies, and interpretations that make up the creation and reception of music videos in general.

The diegetic frame of a pop music video is a complicated matter. While Walther-Hansen's typology may be useful for acousmatic recordings, it is difficult to label any sounds at all in a music video as "meta-" or "extra-diegetic", since the video itself acts to clarify the diegetic role of the sound events who's narrative framing may be in question in the sound recording. In other words, music video confounds the normal conceptualization of sonic diegesis, since unlike other forms of video entertainment, the sound in a music video is arguably the main text while the image serves a supporting role. Thus, the diegetic dichotomy fails to capture the complexities of the sound-image relationship in pop music video. In an essay the explains the use of music video aesthetics in films in general, Vernallis re-iterates that music video is a fundamentally *musical* form:

Free-ranging camera movements like dollying, handheld, reframing, and crane shots reflect music's flowing, processual nature; blocks of image highlight song

structure, intense colourization illuminates features like a song's harmony, sectional divisions and timbre; visual motifs speak to musical ones (…) (2008, p. 277).

If the diegetic dichotomy is, in Kassabian's words "not sufficient to cover the various examples of music that cross over, through, around, and under that boundary" (2013, p. 91), then it is even more so for music videos, who's entire point seems to be to illuminate, demonstrate, exaggerate, and/or complicate the stories told by music recordings. The viewer comes to the music video knowing, in most cases, that it is an extension of an already existing recorded track, and this intertextual duality of the music video implies a multiplicity of entry points to the song's interpretation. Part of what can constitute a song's meaning is in the story it tells, and the viewer's role as the interpreter of musical meaning cannot be ignored, since, contrary to being a passive and external entity, the viewer of a music video is the switch that completes the narrative circuit.

Navigational Agency

At this stage, I want to focus on the construction of 360º music videos. Ultimately, my aim is to show how the experience of the viewer is important in understanding narrative structures in music videos. When used for music production, immersive and interactive media technologies such as VR, surround sound and 3D audio, and 360º videos create a situation where the viewer can be seen as a staged part of the composition (Bresler & Hawkins, 2021). This is because their experiences are centered, and rather than being passive observers of a presented audiovisual scene, they can be placed on the audiovisual stage and thus thrust into an active and participatory role within the music performance. In this section, my goal is to demonstrate how movement implies the possibility for immersion and interactivity. To this end I propose the analytical term *navigational agency*, which describes the immersive pleasures of interacting with the narrative of a music video through spatial movement and control. Navigational agency should be considered as a spectrum, whereby a video can feature varying degrees and modes of it through the types and qualities of movement afforded.

In many cases of 360º music videos, the viewer is granted an easily discernible and defined perspective wherein their placement in the diegesis is explicit enough to be considered a character within the story. Other times, the viewer is placed into the scene as an observer, similar to traditional 2D music videos. In any case, the viewer is invited to participate through interactions with the stage. I suggest that it

is in these interactions that 360º videos and other immersive media formats make explicit an implicit feature of music and music videos in general: that what it means for a song to mean something is an active process that includes the experiences of the viewer. In other words, in pop music, and especially in pop music videos, the construction of diegesis includes the viewing experience itself. As the user interacts with the music video through various means, the participate in creating the very narrative they consume.

This notion is supported through the concept of ecological perception, an idea first introduced by James Gibson in psychology (Gibson, 1977, 2015) and brought into musicology and music psychology through Clarke's notion of an ecological approach to the perception of musical meaning (Clarke, 2005). Clarke posits that meaning comes forth from the confluence of the listening "environment" (a technical term which encompasses not only the space and place of the listening but also the background, taste, and experience of the listener) and the musical performance, be it recorded or live, replete with its various structural affordances (ibid.). Importantly, while the term "affordance" implies a kind of structuralism where particular structures in the pop score demand particular responses from listeners, Gibson maintains that affordances have a dialectical quality that imply "the complementarity of the animal and the environment" (Gibson, 2015, p. 119).

Writing about digital hypertext narratives, Murray reminds us that "activity alone is not agency" (2016, p. 124). She suggests that agency is more than the sum of the interactive participations of the viewer and defines it as the "satisfying power to take meaningful action and see the results of our decisions" (2016, p. 123). This definition is useful, but unclear is what constitutes a "meaningful action." Is it necessary that the user can do whatever they want without restriction? Or can the medium place constraints, even large ones, on the viewer's possible actions while still yielding degrees of agency? I argue that the freedom to navigate space is, on its own, 'meaningful', in the sense that it opens up the possibility for new kinds of meaning.

In general, music videos are not hypertexts—viewers do not make decisions that constitute direction for the narrative and, regardless of the viewer's actions, the plot will unfold in the same way. However, immersive music videos offer the viewer interactivity in the form of spatial navigation, where the narrative unfolds around the viewer, and she must use her body to actively engage with the text to experience it fully. In this way, while immersive music video is not hyper textual, it can nonetheless be considered a form of ergodic cybertext, where "nontrivial

effort is required to allow the reader to traverse the text" (Aarseth, 1997, p. 1). Although music videos are, by definition, linear in the sense that they follow musical form, interaction with the virtual space allows the reader to be "constantly reminded of inaccessible strategies and paths not taken, voices not heard" (Aarseth, 1997, p. 3). Indeed, Murray reminds that spatial navigation is itself a highly pleasurable form of interactivity: "construing space and moving through it in an exploratory way… is a satisfying activity regardless of whether the space is real or virtual (2016, p. 125).

In dealing with navigational agency in VR and 360º music videos, there are several spatial and navigational aspects to consider:

> *Stage configuration* – what is the size, shape, and depth of the virtual environment?
> *Degrees of freedom* – how much movement is the viewer afforded?
> *Range of motion* – what are the limits of this movement?

Stage Configuration

The ideal implementation of virtual reality is normally imagined as something like the Holodeck: the fictional room from "Star Trek" in which the user enters, tells the computer the parameters of the environment and story they would like to experience, and the room transforms to the specifications, creating for the user a completely accurate sensory experience that is indistinguishable from the real thing. While, of course, the analogy of the Holodeck has been impossible to actually deliver, this kind of imagination for what VR could someday be has driven much of the research and interest into VR since its inception in the 1990s (Murray, 2016). Spatially, the holodeck metaphor demonstrates what is ultimately necessary to create a virtual spatial environment. Marie-Laure Ryan claims that "being inside a computer-generated world involves three distinct components: a sense of being surrounded, a sense of depth, and the possession of a roving point of view" (Ryan, 2001, p. 53).

Through degrees of freedom and range of motion, concepts I address in the next sections, immersive music videos feature this "roving point of view". Perceived dimensions of the environment and their morphologies are a useful starting point for discussion. In an earlier study we conducted, Hawkins and I devised a model of Virtual Audiovisual Space (VAVS) for describing and interpreting the spatial configuration of the VR audiovisual stage (Bresler &

Hawkins, 2021). Inspired by Trevor Wishart's model of Virtual Acoustic Space (Wishart, 1985), this model uses as its basis Camilleri's model of sonic space to describe the position, disposition, and temporal unfolding of sound and visual objects (Camilleri, 2010). In addition it builds on Denis Smalley's notion of source-bonding to describe the connection of sounds within a scene to supposed causes (Smalley, 2007), and our own notion of the aesthetic space, which comprises the interpreted meanings that are synthesized in the viewer's audiovisual experiences. As Hawkins and I have argued, any interpretation of space within this context necessitates a description of the apparent size, shape, and quality of that space.

Sound and image are not always aligned in their spatial construction. What I mean is that in music videos, while the video may be shot in a real space (an outdoor stage, a small room, in a warehouse) the sounds of the pop recording are not altered to match the expected sonic properties of the video's scenic space. This observation is quite obvious, but it is worth mentioning since 360º music videos often feature 3D visuals with static, stereophonic audio. In other words, while the viewer is invited to interact and move through the visual space, the sonic space remains a fixed stereo image that, depending on whether viewing in a head-mounted display or on a screen at a distance, either follows the user, or seems to be an unmoving element. Moreover, just like in 2D music videos, 3D videos commonly feature animated visual scenes without analog in the physical world and for which the viewer would have no auditory reference for the acoustic properties the sound in such a space should have.

For example, in the video for Muse's "Revolt", the scene opens with the sounds of sirens and driving cars while text overlays the screen explaining banishment of "freedom" in 2025 as "government drones fill the sky." A moment later, the police vehicles appear in frame, black, tank-like SUVs which stop to release masked officers. Panning around the view in the 360º video, it is noticeable that the sounds do not move to reflect the viewer movements—the police car sirens to not appear to come from the direction of the cars themselves, rather they are in stereo as it is in the acousmatic recording. Shortly after the band begins performing outside in the open, positioned as if on stage at a rock concert while their audience is the clashing of military police and rebellious protesters. What is heard over the video is in fact the stereo release of "Revolt," without any spatial change to reflect the outdoor scene or any spatialization to position the band's performers in the direction they are with respect the viewer's facing in the 3D virtual space.

Figure 2: "Revolt" by Muse: view of bassist Chris Wolstenholme with the riot happening in the background. Visible in frame is a "government drone."

These contradictory spatialities are not surprising, given that they are part and parcel of the music video paradigm. Brøvig-Hanssen and Danielsen confirm that these surreal spatial configurations have "a tendency to point the listener toward a real-world physical phenomenon even as it acts to undermine that reality" (2016, p. 27). In the above example, the two physical phenomena are those of the night-time, outdoor protest clash and the indoor, studio-polished recording of the rock band Muse. As these two realities fade into each other they do not cause conflict for the viewer. On the contrary, they work together abstractly to communicate an effective narrative statement on civil conflict. Coming from compositional studies, Smalley has referred to "spatial simultaneity" as the phenomenon where the listener can be, without conflict, "aware of simultaneous spaces" that are either implicit or explicit, or where "the listener remains aware of the existence of a space in its absence" (1997, p. 124). While this refers to the multiplicity of spatialities within audio recordings, I argue that it is just as well to consider the contradictions in spatiality between sound and image (or in many cases, both within the recording, within the image, and in the sound and image) in terms of spatial simultaneity.

Degrees of Freedom

In the field of immersive and interactive media technologies, the level of designed spatial control is often described in terms of degrees of freedom (DOF), where each degree represents a possible axis for movement in the virtual space. It is

128

important to note that the use of the term "freedom" here is entirely technical and does not correspond to any social or political notion on freedom, rather, it should be understood in comparison to media which do not offer the user movement through the space of the video (such as a traditional 2D music video). The first three axes, which comprise the total movement possibilities for so-called 3DOF media, are the rotational axes, wherein the user can move within the parameters of yaw (swivel the head as in the "no" gesture), pitch (as in nodding "yes"), and roll (tilting the head towards the shoulders). These dimensions combine to represent all the possible movements of the head at the level of the neck without moving the shoulders, and technologically they are simple to implement since when recording 360º videos, the 360º camera is a stationary object that records a 3D visual field. Thus, the 3DOF video allows the viewer to rotate their relatively narrower frame of view within this stationary spherical or cylindrical image.

In animated or digitally manipulated videos, three more axes can be added to create the 6DOF environment, where the viewer can additionally move along the translational axes: forward-backward, left-right, and up-down. Importantly, this is impossible with live-action video recordings since one cannot place a camera at all the possible points in a scene where a viewer may want to position themselves. Regardless, much research has been done on producing 6DOF audio recordings, for example using combined third-order ambisonic microphones (Rivas Méndez, Armstrong, Stubbs, Stiles, & Kearney, 2018). While these kinds of techniques for recording make 6DOF sound possible, mono and stereo recordings mixed in virtual 3D sound formats such as ambisonics or Dolby Atmos seem to be the preferred method for recordists attempting to produce content for 3D systems.

Figure 3: "The Hills remix" by The Weeknd feat. Eminem: front and back views as meteors destroy the city.

In many cases, producers of live-action 360º videos will move the camera during capture, thus creating translational movement for the viewer, although they only have rotational control. Currently, most immersive music videos, and all the music videos discussed in this essay, are 3DOF live-action or animated videos.[3] For example, in The Weeknd's "The Hills remix," the viewer is granted a perspective of a hovering camera that follows the artist's slow walking out into a dystopian urban street at night (figure 3). Initially, the view faces The Weeknd directly, although he never gazes to the camera, instead staring straight ahead with a look of complete absence. As the viewer rotates, they can see what looks like asteroids crashing down on the city.

Looking back at the artist for reaction, we see none, and instead continue the slow progression through the scene. Importantly, this clip shows how rotational movement combines with pre-defined lateral movement to create a forward sense of motion. The video moves with the artist who is walking in a slow procession, but the camera movements are not steady. Instead, they bob up and down with each step, creating the impression that we too are walking with the artist, or perhaps that

---

[3] In general, 6DOF experiences currently require a complete virtual reality implementation such as an Oculus Rift headset, since the CPU processing required for playback is significant and cannot be effectively streamed or played on most mobile devices. A striking example of this is Björk's *Vulnicura VR*, an album of VR music videos available on the PC gaming shop Steam. The video from this album has been researched with Stan Hawkins in another study.

we are his own out-of-body experience in this jarring scene of destruction. Regardless of the amount of movement afforded to the viewer, the existence of any "degrees of freedom" at all provides the user with navigational agency, especially when there exists visual or auditory content that is only accessible through re-orientation (a concept to which we will return). Although the narrative will unfold similarly in each viewing, navigational agency allows the viewer to interact with the scene and place themselves within it.

Range of Motion

Even within the freedom of motion granted in an immersive video, there exists variation in the real and perceived range of that motion. For example, although we have only discussed 360º video so far, the producer can also choose to limit the possible field to a 180º frame for various reasons. For one, cutting the field of view in half, effectively into a single hemisphere, can allow for a higher fidelity image since a smaller frame can allow for higher resolutions. Additionally, this choice can set limits on navigational agency by design, since the producer may feel they need to restrict the movements of the viewer into this single hemisphere. A definition of agency such as that I cited earlier by Murray (Murray, 2016, p. 124) can imply that part of what enables it in cybertext is a one-to-one relation between a taken action and an intended result. Arguing that such relation can be difficult to define, Mason has suggested that movement alone does not constitute diegetic agency, instead suggesting it to be affect, which "is a necessary path to agency… and we must be fluent with our means of affect to experience immersion" (Mason, 2013, p. 31).

This is in agreement with my anecdotal experience of viewing an immersive video for the first time. It seems natural to begin by searching for the boundaries of the experience, asking myself: is the video a full 360º or 180º? Does my view have an avatar or a body? Do I have the freedom to move in space, or simply rotate my perspective? In practice, this means that the first experience of a VR or 360º production is, for me, one of determining the range of motion possible, and this activity of finding the boundaries serves to increase the chances of my having an immersive experience, since what is required is that I feel able to make free and meaningful choices. By knowing the limitations of my choices, I can more easily make the kinds of choices that are possible. Without taking this step, a viewer might, in the middle of a video, suddenly decide to turn around and find they cannot, or try to move their hand and find they do not have one. Any of these types

of experiences only serves to remind the viewer of the non-reality of their experience, ultimately taking them out of it.

Diegetic Immersion

Immersion, in general, is the experience of losing oneself in within an activity and it is often likened to experiences such as Csikszentmihalyi's notion of the flow state (1990). Thus, immersion can be described in terms of the pleasure of a repetitive action—the feeling of losing time when engage in enjoyable, repetitive, and comprehendible activities. But it can also occur, as in the flow state, in activities that are the right amount of both challenging and engaging, such as the cognitive task of reading and comprehending a difficult text, since this experience can be to the reader both profound and empowering.

Notions of immersion and "letting go" have also been a central part of studies of dance club music and club culture, and this is fundamentally tied to notions of temporality. For example, Frith has insisted that "dance is not just to experience music as time, it is also to experience time as music… more intense, more interesting, more pleasurable than 'real' time" (1996, p. 156). Similarly, Hawkins has noted how the dance floor can enable "the sensation of being 'loved up' (an expression often used by DJs and clubbers) [which] suggests a state where the body of the individual or the crowd is immersed in sound" (2008, p. 122). Importantly, these immersive club experiences are enabled by musical features such as the beat and the groove, as well as environmental and social factors such as the lighting, volume, and involvement of the crowd.

As I have implied so far, immersion is often pitted as the antecedent to agency in multimedia like games and hypertext narratives, since high degrees of agency are seen as breaking the story into small, difficult to synthesize parts, while higher degrees of immersion require more complex narratives that require consistency and reduce the possibilities for agency (Ryan, 2001). So, what constitutes the elements that form diegetic immersion within 360º videos? I argue that there are two main factors that dictate the propensity for diegetic immersion. The first is in the construction of the relevant visual field, what I am calling *visual saturation*. Second is the *perceived role*, which is correlated to the viewer's narrative embeddedness and designed experiences of embodiment.

Visual Saturation

Another aspect of visual spatiality in 3D media is viewer mobility. Although the viewer may be free to move within a visual scene, it is not necessarily the case that there is something "happening" in every part of the scene. The amount of space containing engaging visual material needs to be considered. Albeit, what may be considered "engaging" or "interesting" in this context is certainly a matter for individual interpretation, since the absence of image can be just as engaging as the presence of one. Still, it is true that in music videos, directors guide the viewer's gaze through camera movements, framing, color, and lighting in order to invite them into diegetic immersion. Here I offer the analytic term *visual saturation*, which refers to the amount of utilized space within the visual field and how the producers of the video have used these visual features to suggest and guide the viewer. Importantly, visual saturation is different from the spatial construction of the 360º or 180º video, which I have already discussed. In that sense, one can talk about the perceived size and shape of the stage in terms of what is possible. Here, I refer to something more qualitative and hermeneutic, which is the amount of the visual field that the viewer finds relevant to explore in their viewing experience.

An example from the opening of Taryn Southern's "Life Support" illustrates this well. The scene begins in a spooky wood as the viewer is moved towards a lone, run-down house with a nighttime city skyline visible in the distance. Once the viewer arrives at the house, however, they are transported into a vast, dark space with nothing surrounding them except a large, strange machine flanked on both sides by rectangular screens showing images of brain scans (figure 4). Turning around, the viewer will see that this machine apparatus is, for some time, the only visible object in the entire video, which is just as well because it is visually captivating with its moving arm waving around a human body like a rag doll. After some time, lights and images begin to emanate from the machine, moving past the viewer to the rear of the scene, drawing their attention there to notice that there are now things happening behind them—flashing and moving light patterns that reveal parts of a seemingly infinite darkness.
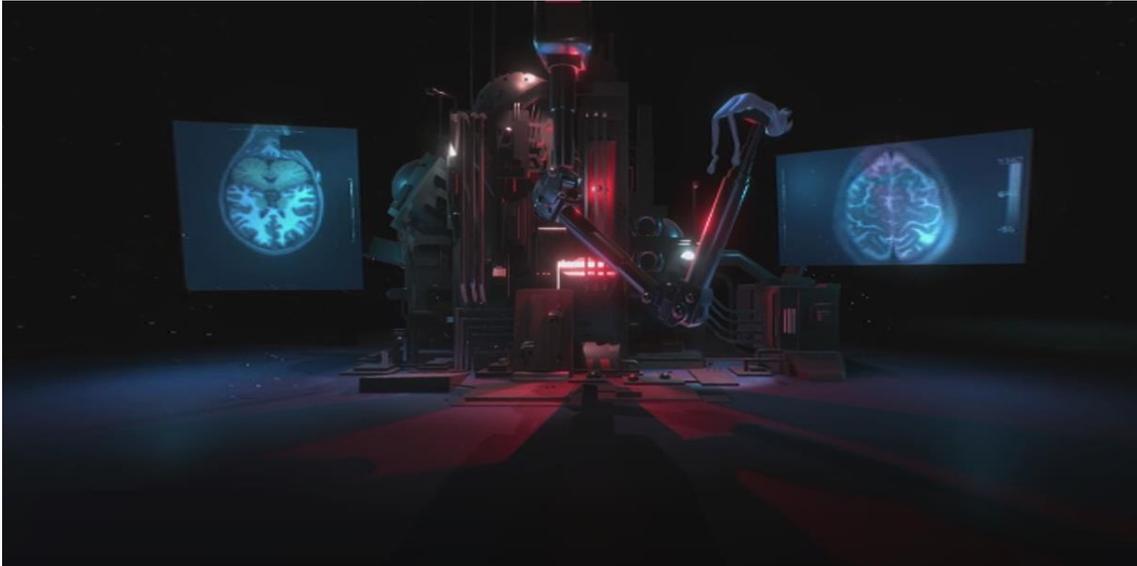
Figure 4: "Life Support" by Taryn Southern: a large and mysterious machine interfaces with a lifeless (for now) humanoid body.

Here, the producers of the video have carefully crafted the viewer's visual attention by first revealing a space they can freely explore (the woods) before drawing their attention to a narrow frame (the machine) which encourages them to remain still for a moment. Finally, they slowly open up the visual scene with moving lights, reminding the viewer of the immersive qualities of their visual experience. These strategies serve to invite the viewer to explore the space while guiding them towards the most relevant visual aspects of the scene.

Perceived role and viewer subjectivity

In thinking about how immersion functions within so-called immersive audiovisual media, it is important to ask how embedded into the story is the viewer. Because the video is 360º, there will be audiovisual material which surrounds the listener, who will have some degree of navigational agency within the space. However, their narrative embeddedness is a question of their role within the story. In short, the answer to this question will lie somewhere on a line between an outside observer and independent narrative agent. In other words, is the viewer a passive observer or an active participant, and to what degree?

At this point I turn my attention to the notion of the subject position in film studies, which Johnston has defined as "the way in which a film solicits, demands even, a certain closely circumscribed reading from a viewer by means of its own formal operations" (Johnston, 1999, p. 333). In other words, its use is an attempt to allow for analyses of meaning that are constructed both in the formalization of

the film and its reception, thus skirting the failures of both structuralism and post-modernist relativism. The use of subject position to describe the listener's role in popular music meaning has been used by several. For example, Clarke has suggested that the listener's own subject position "results from the separation between the narrative content of the film and the manner in which viewers are allowed, or invited, to know about that narrative" (2005, p. 93), and suggesting in music that the narrative content and the framing of subject position occurs "not through the semiotic language of 'codification,' but through the perceptual principal of 'specification'" (ibid. pg. 125). In other words, although the viewer is ultimately the arbiter of meaning, their interpretations are nonetheless shaped in part by structural elements of the music which specify relationships and correlate to particular responses. I do not entirely concur with Clarke—the semiotic language of audiovisual codes is but one way of explaining the structures that lend themselves to perceptual specification. In music analysis, one can only accurately explain such textual elements, one's own interpretation of them, and perhaps some alternative interpretations they can imagine.

How then does immersive media elicit subject positions differently from traditional films, recorded tracks, and music videos? Here I would like to extend on the idea of subject-position by suggesting that immersive media engage directly in subject-position*ing*. That is, through the staging of the viewer directly on the stage, and in particular through their freedom of movement on the stage, the 360º video has become a platform for the viewer to participate in positioning their own subjectivity in the audiovisual scene. Granted, this embeddedness can be aided through implication of a character role.

Demonstrating this by example, I return to Muse's "Revolt." At first the viewer seems to be simply an observer—they are moved through the scene which was presumably recorded on a moving 360º video camera, going back and forth between the protest and the performing band. Embedded within the visual field are futuristic, digital circular overlays which seem to identify objects within the scene such as a person's face or a vehicle in the background, displaying illegible data on the various identified objects, reminiscent of the first-person views from films like The Terminator and Robocop (figure 5). In the opening on-screen text, we were told of "government drones," which are visible floating around the scene, and it quickly becomes clear to the viewer that their perspective is that of one of these autonomous, robotic surveillance cameras. Through clever manipulation of the video, the producers have placed the viewer firmly within the diegesis, giving them

the privileged view of the imagined government overseers. The band also stage themselves as sympathetic to such causes, as the cameras hover over the musicians they identify their faces in the same way as those of the revolting citizens. While the viewer has no control over their lateral movements, they nonetheless have rotational control, and looking around at various people and objects as they are automatically scanned and identified, one cannot help but feel a sense of complicity.

Figure 5: "Revolt" by Muse: a protester has been identified by the viewer's drone with the text "Target Armed."

Assisting in the development of the user-character's role is the presence of a body or an avatar. The above example illustrates this to an extent—the embodiment in first-person of a surveillance drone is confirmed through the overlaid surveillance data, and the erratic movements of the camera, which mimic the movements of the other drones that are visible in the video, encourage the user to take the role of the camera by rotating their own view in erratic ways. Going further, one can be granted a human body or bipedal avatar, which can serve as a stand-in for one's own body and heighten the embodied experience. For instance, in Squarepusher's "Stor Eiglass," the viewer finds themselves in a neon, psychedelic dreamscape, moving steadily through a barrage of imagery that conjures up memories of video games, 1980s shopping centers, and the imagined sci-fi city of the future (but in high-contrast). When the viewer looks around in the scene, they will notice upon looking down that they have been given a body (figure 6)—naked and cartoonishly

shaped, and with a cleanly severed neck (complete with a visible bone) just below the point of view, as if the viewer's head is floating above.
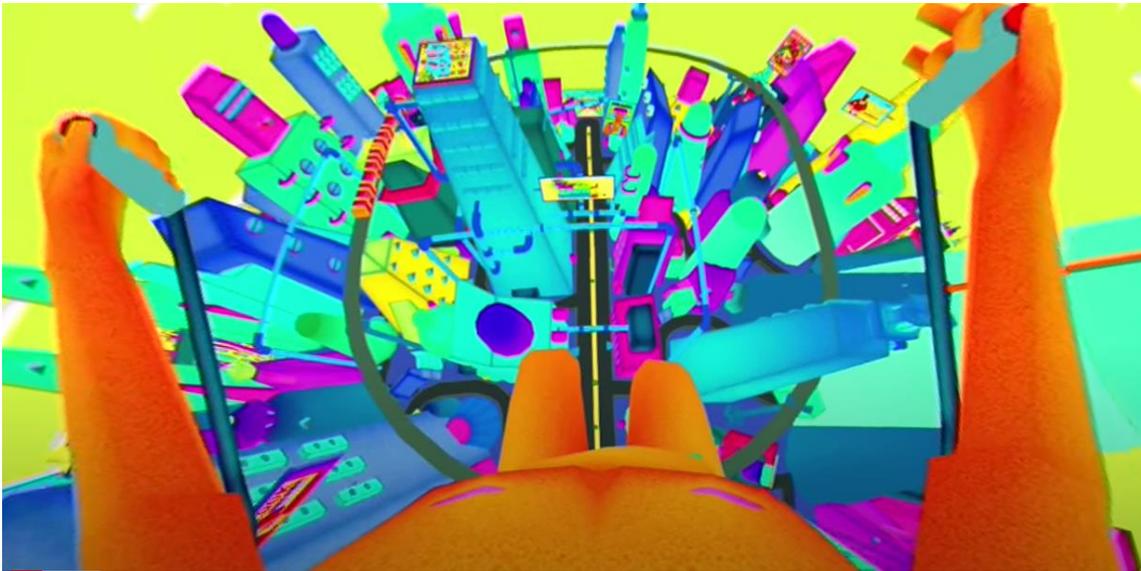

Figure 6: "Stor Eiglass" by Squarepusher: Looking down at the psychedelic city, the viewer's "body" is visible.

The body appears to be sitting down at first with arms extended and gripping a set of joysticks, but then throughout the vehicle on which we travel changes, becoming at one point a bicycle, and eventually it goes away, and we see our character walking. As the song progresses, the body changes, at one point suddenly becoming a woman, with large, naked breasts now obfuscating some of the view below. And later in the video, as the song gets more energetic and the imagery becomes more and more psychedelic and fractal, the body disappears entirely as the viewer finds themselves in an overwhelming, symmetrical, spinning scene of changing shapes and colors.

The above example with its realistic naked body demonstrates very clearly how the gendered body is always part of the design of embodied experiences. In other words, whenever media imply an embodied experience or subject position, it is critical to ask the question: whose body is it that is being implied? A major part of the utopian ideology of the digital-virtual environment is the freedom that the digital world grants us in transforming our "creative thoughts and imagination" into "reality and actuality through digital means" (Rambarran, 2021, p. 1). The example of "Stor Eiglass" depicts a parodic spin on this, as the nude body on display is at first coded male, and later (and without warning) female—at times it is completely motionless, and other times it moves in an autonomic fashion.

Always visible is the decapitated neck upon which the viewer's lens resides. Ultimately, Squarepusher offers a humorous critique of this utopian virtual ideology through this imagery, illustrating that in media that purports to transform a person into their ideal digital selves, the best it can offer is a new set of interchangeable avatar categorizations.

Conclusion

Being immersed in a story is a fundamentally human experience, and thus it is no surprise that the multitude of technologies for multimedia storytelling are so concerned with assisting us to more easily find such experiences. While the discourses around immersion in film, music, video games, and other forms of media often focus on the distinction between agency and immersion, I have attempted here to convince you that both agency and immersion are in fact allies in storytelling. In different forms of media, they function in different ways. For example, in video games, players have much higher degrees of agency than in more structured media such as film, but there still exists a wide range of agency from the auto-scrolling single-control interactivity of mobile games like "Flappy Bird" to the total open-world possibilities in games like "Zelda: Breath of the Wild" (Collins, 2013). Within this range there are many levels of complexity within the stories that are told, or able to be told. The same is true for other media—the introduction of expanded modes of access and interaction create a different range of possibilities for storytelling.

Music videos are a special form of media. Kelly has insisted that they are "always already a hybrid medium, comprising audio and visual forms and structures that intersect and interrelate in ways that can be described as intermedial" (Kelly, 2019, p. 219). Unlike other forms of film, television, or video, where music extends the interpretive possibilities of the visual and dialogic narrative, music videos do the opposite, using visuality to extend the hermeneutic position of the musical text. Considering popular music in new, immersive and interactive forms of media, including 360º videos, give analysts recourse to consider anew the ways that subject positioning can occur in pop multimedia. The formation of pop music video diegesis, I have shown, is not only a structural and musical phenomenon, but it is itself dialogical. In other words, viewers of music videos are the co-creators of narrative structure. I believe that 360º music videos offer an easy-to-demonstrate case for this, since the way they stage listeners within the story world is obvious. However, I would conclude by suggesting that these

processes are not exclusive to music presented in these technologically innovative ways. Viewers of music videos have always been nexus of audiovisual meaning and while the story is told by the creators of a music video, the diegetic frame is only complete when we acknowledge the role of the viewer in its formation.

References:

Aarseth, E. J. (1997). *Cyberext: Perspectives on Ergodic Literature*. Baltimore: Johns Hopkins University Press.

Ålvik, J. M. B. (2017). "Armed with the faith of a child": Marit Larsen and strategies of faking. In S. Hawkins (Ed.), *The Routledge Research Companion to Popular Music and Gender* (pp. 253-266). New York: Routledge.

Auslander, P. (2008). *Liveness: Performance in a Mediatized Culture* (2nd ed.). New York: Routledge.

Auslander, P. (2021). *In Concert: Performing Musical Persona*. Ann Arbor, MI: University of Michigan Press.

Bresler, Z. (2021). Immersed in Pop: 3D Music, Subject Positioning, and Compositional Design in The Weeknd's 'Blinding Lights' in Dolby Atmos. *Journal of Popular Music Studies, 33*(3).

Bresler, Z., & Hawkins, S. (2021). [Forthcoming] "A Swarm of Sound": VR immersion in Björk's video 'Family'.

Brøvig-Hanssen, R., & Danielsen, A. (2016). *Digital Signatures: The Impact of Digitization on Popular Music Sound*. Cambridge: MIT Press.

Burns, L. (2016). The Concept Album as Visual-Sonic-Textual Spectacle: The Transmedial Storyworld of Coldplay's Mylo Xyloto. *IASPM Journal, 6*(2), 91-116.

Burns, L., & Hawkins, S. (Eds.). (2019). *The Bloomsbury Handbook of Popular Music Video Analysis*. New York: Bloomsbury.

Burns, L., & Woods, A. (2019). Humor in the "Booty Video": Female Artists Talk Back Through the Hip-Hop Intertext. In T. M. Kitts & N. Baxter-Moore (Eds.), *The Routledge Companion to Popular Music and Humor*. New York: Routledge.

Camilleri, L. (2010). Shaping sounds, shaping spaces. *Popular Music, 29*(2), 199-211.

Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. New York: Oxford University Press.

Collins, K. (2013). *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. Cambridge: MIT Press.

Csikszentmihalyi, M. (1990). *Flow. The Psychology of Optimal Experience*. New York: Harper Perennial.

Darley, A. (2000). *Visual Digital Culture: Surface place and spectacle in new media genres*. London: Routledge.

Dibben, N. (2013). Visualizing the App Album with Björk's Biophilia. In C. Vernallis, J. Richardson, & A. Herzog (Eds.), *The Oxford Handbook of Sound and Image in Digital Media* (pp. 682-704). New York: Oxford University Press.

Frith, S. (1996). *Performing Rites: On the Value of Popular Music*. Cambridge, MA: Harvard University Press.

Gibson, J. J. (1977). The Theory of Affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting and Knowing: Toward and Ecological Psycology*. Mahwah, NJ: Lawrence Erlbaum.

Gibson, J. J. (2015). *The Ecological Approach to Visual Perception* (3rd ed.). New York: Psychology Press.

Gorbman, C. (1980). Narrative Film Music. *Yale French Studies, 60*, 183-203. doi:10.2307/2930011

Hansen, K. A. (2019). (Re)Reading Pop Personae: A Transmedial Approach to Studying the Multiple Construction of Artist Identities. *Twentieth-Century Music, 16*(3), 501-529. doi:10.1017/S1478572219000276

Hawkins, S. (2002). *Settling the pop score: Pop texts and identity politics*. Burlington, VT: Ashgate.

Hawkins, S. (2008). Temporal Turntables: On Temporality and Corporeality in Dance Culture. In S. Baur, J. Warwick, & R. Knapp (Eds.), *Musicological Identities: Essays in Honor of Susan McClary* (pp. 121-134). New York: Routledge.

Hawkins, S. (2009). *The British pop dandy: masculinity, popular music and culture*. New York: Routledge.

Hawkins, S. (2016). *Queerness in Pop Music: Aesthetics, Gender Norms, and Temporality*. New York: Routledge.

Jirsa, T., & Korsgaard, M. B. (2019). The Music Video in Transformation: Notes on a Hybrid Audiovisual Configuration. *Music, Sound, and the Moving Image, 13*(2), 111-122.

Johnston, S. (1999). Structuralism and its Aftermath. In P. Cook & M. Bernink (Eds.), *The Cinema Book* (2nd ed., pp. 323-341). London: British Film Institute.

Kassabian, A. (2013). The end of diegesis as we know it? In J. Richardson, C. Gorbman, & C. Vernallis (Eds.), *The Oxford Handbook of New Audiovisual Aesthetics*. Oxford: Oxford University Press.

Kelly, J. (2019). The Palimpsestic Pop Music Video. In L. Burns & S. Hawkins (Eds.), *The Bloomsbury Handbook of Popular Music Video Analysis* (pp. 219-233). New York: Bloomsbury.

Korsgaard, M. B. (2013). Music Video Transformed. In J. Richardson, C. Gorbman, & C. Vernallis (Eds.), *The Oxford Handbook of New Audiovisual Aesthetics* (pp. 501-524). Oxford: Oxford University Press.

Korsgaard, M. B. (2017). *Music Video After MTV: Audiovisual Studies, New Media, and Popular Music*. New York: Routledge.

Korsgaard, M. B. (2019a). Changing Dynamics and Diversity in Music Video Production and Distribution. In L. Burns & S. Hawkins (Eds.), *The Bloomsbury Handbook of Popular Music Video Analysis* (pp. 13-26). New York: Bloomsbury.

Korsgaard, M. B. (2019b). SOPHIE's 'Faceshopping' as (Anti-)Lyric Video. *Music, Sound, and the Moving Image, 13*(2), 209-230.

LaFrance, M. (2013). Celebrity, Spectacle, and Surveillance: Understanding Lady Gaga's 'Paparazzi' and 'Telephone' through Music, Image, and Movement In M. Iddon & M. L. Marshall (Eds.), *Lady Gaga and Popular Music*. New York: Routledge.

Liljedahl, A. A. (2019). Musical Pathfinding; or How to Listen to Interactive Music Video. *Music, Sound, and the Moving Image, 13*(2), 165-185. doi:https://doi.org/10.3828/msmi.2019.10

Mason, S. (2013). *On Games and Links: Extending the Vocabulary of Agency and Immersion in Interactive Narratives.* Paper presented at the ICIDS 2013, London.

Morie, J. F. (2007). Performing in (virtual) spaces: Embodiment and being in virtual environments. *International Journal of Performance Arts and Digital Media, 3*, 123-138. doi:10.1386/padm.3.2-3.123_1

Murray, J. H. (2016). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace* (2 ed.). New York: The Free Press.

Rambarran, S. (2021). *Virtual Music: Sound, Music, and Image in the Digital Era*. New York: Bloomsbury Academic.

Rivas Méndez, D., Armstrong, C., Stubbs, J., Stiles, M., & Kearney, G. (2018). *Practical Recording Techniques for Music Production with Six-Degrees of Freedom Virtual Reality*. Paper presented at the 145th Audio Engineering Society Convention, New York.

Ryan, M.-L. (2001). Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media. In. Baltimore: Johns Hopkins University Press.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised Sound, 2*(2), 107-126.

Smalley, D. (2007). Space-form and the acousmatic image. *Organised Sound, 12*(1), 35-58.

Vernallis, C. (2004). *Experiencing Music Video: Aesthetics and Cultural Context*. New York: Columbia University Press.

Vernallis, C. (2008). Music video, songs, sound: experience, technique and emotion in Eternal Sunshine of the Spotless Mind. *Screen, 49*(3), 277-297.

Vernallis, C., Herzog, A., & Richardson, J. (Eds.). (2013). *The Oxford Handbook of Sound and Image in Digital Media*. Oxford: Oxford University Press.

Walther-Hansen, M. (2015). Sound Events, Spatiality and Diegesis – The Creation of Sonic Narratives in Music Productions. *Danish Musicology Online*, 29-46.

Winters, B. (2010). The non-diegetic fallacy: Film, music, and narrative space. *Music and Letters, 91*(2), 224-244. doi:10.1093/ml/gcq019

Wishart, T. (1985). *On sonic art* (2nd ed.). Amsterdam: Routledge.

Wood, A. (2002). Timespaces in spectacular cinema: crossing the great divide of spectacle versus narrative. *Screen, 43*(4), 370-386.

Article 4 – "Hope to Die": Compositional Design and Queer Subjectivity in the Music Videos of Orville Peck

Zack Bresler and Stan Hawkins

Chapter is accepted and in editorial review for an international anthology at time of submission, to be published in 2022

Introduction

The enigmatic country-pop artist, Orville Peck, would produce some of the most awe-inspiring sound recordings during the second decade of the twenty-first century. Born somewhere in the Southern Hemisphere, around 1987/1988,[1] he has been fastidiously elusive about his origins and personal life, and known for covering up his face with elaborate masks.[2] From the little he has revealed, he is son of a sound engineer and spent a good deal of his childhood doing voice-overs for cartoons. He also trained as a ballet dancer for twelve years during his youth, studied acting at The Royal Academy of Music and Dramatic Art in London, and took part in West End musicals.[3] Residing in Canada, at the time of conducting this research, Peck has established an international career as a queer country singer, always confirming his gay sexuality. So large is his following that on the 50th anniversary of the LGBTQ Pride events in 2020, *Queerty Magazine* would rank him amongst the top fifty heroes who have fought for liberty, dignity, and acceptance for all people.[4]

During the course of this chapter, we concentrate on the correlation between sound production, imagery, and subjectivity in the track 'Hope to Die', from

---

[1] As revealed in an interview with *L'Officiel Magazine* in March 2019, on the day of the release of the record *Pony*, on which the song we analyze in this piece is found: https://www.lofficielusa.com/music/orville-peck-interview-2019.

[2] While many have speculated about Peck's real name, he has been insistent in maintaining his privacy, saying in a statement in *The Guardian* "there is a temptation to try and unmask what I do, but to do so would be to miss the point entirely." In the previously referenced interview with *L'Officiel*, he has stressed his study of "mask as an art form… the method made famous by Jacques Laqoc." While speculation surrounding his 'true' identity is rife on the internet, we choose to not engage with it explicitly in this essay. https://www.theguardian.com/culture/2019/nov/19/orville-peck-i-grew-up-feeling-alienated-so-i-became-a-lone-cowboy.

[3] This biographical information about his parents, education, and West End experience was revealed in an interview on the podcast "Sloppy Seconds with Big Dipper and Meatball" on May 1, 2020: https://foreverdogpodcasts.com/podcasts/sloppy-seconds/.

[4] https://www.queerty.com/pride50/

Peck's debut album *Pony*, released in 2019. Notably, Peck played all the instruments himself and had a major say in the production. While the predominant stylistic trait is standard country music, it is mashed up with numerous other references. Our prime purpose is to examine the sonic details of production and issues of audiovisual representation, and our focus therefore falls on the official track and video of 'Hope to Die'. The analytic methods employed seek to foreground the aesthetic effects of production, audio engineering, and compositional structure. They are also intended to uncover the congruences between technologies of music production and performativity; sonic devices such as reverb, delay, stereo panning, balance, vocal compression, and instrumentation immerse the listener/viewer serve to resignify country music. By harnessing an Americana stylistic aesthetic, Peck's digital marker comprises a range of stylistic and technical codes, with influences of artists, such as Chris Isaac, Elvis Presley, Dolly Parton, Steven Morrissey, Whitney Houston, Prince and others, clearly evident. As such, a range of innovative technologies are employed as part of a process of resignification. In our analyses we adhere to approaches that highlight strategies of listening that give way to a critical evaluation of a song's recording, maintaining that meaning results from both observation and evaluation. Peck's politics of representation escort us on a journey that starts with a discussion of the compositional design, elements of recording and production, and culminates with a consideration of his visual performance.

Compositional design – stylistic and technical coding

Musical features in the track 'Hope to Die' disclose what we define as the *pop score*, namely the totality of the sound recording and all the complexities that are invested in engineering and production.[5] It includes a conglomeration of musical codes that entice the listener into a process of cognizing, enjoying, and relating to artist or band in question. The pop score's design and its 'syntagmatic primary codes' can be broadly categorized as stylistic and technical.[6] It is these that shape the design of a track structurally, mediating the experience of artistic expression. Engaging with an integration of the compositional elements – structure, space, rhythm, timbre, and production – we set out to explore how they constitute the

---

[5] By *pop score* we refer to Stan Hawkins theory of the recorded format of the pop song. See Hawkins 2002.
[6] See Hawkins 2002, 10.

recording. Our reflection of sonic material in its entirety aspires to what William Moylan identifies as the 'qualities and subtleties of recorded sound' that 'pull the listener into understanding and perceiving recorded sounds for their unique individual characteristics, relationships, and the sound qualities they form when combined' (Moylan 2020, 191). Illustrations, transcriptions and typologies are devised to highlight the unique attributes in 'Hope to Die', where the focus on country as a genre is articulated through the syntax of recorded elements that function according to listener perception and competence. Perhaps the most critical task is to consider how recorded sounds function in a way that proffers "insights into the ways in which musical codes are manipulated to create expression through invocations of resistance, compliance, and pleasure" (Hawkins 2002, 12).

The recording under scrutiny raises important questions of vocal presence and authority that are charged for drawing the listener into a specific space.[7] For instance, during the verse, the vocal part establishes the mood and in our reading the sentiments of heartbreak through a sense of 'intimate spatiality'. By this we are referring to the minute characteristics of vocal track that extract a sense of mournfulness. Heavily processed in the mix, Peck's low bass register is ornamented with devices, such as portamento, vibrato, and a subtle use of compression that heightens the physiology of the vocal folds; the shaping of mouth and throat in recording and microphone techniques magnify a wealth of details. On closer inspection, it is also his vibrato that contributes to timbral coloration; the sobbing, trembling quality at the end of most of the phrases on long sustained pitches exemplifies this well. Frequently his voice comes across lonesome through its isolation rest of the recording. The spectrograph (Figure 1) illustrates the slow and wide vocal vibrato on the words 'way' and 'were' at the end of the first line. Evident here is the intensity and the strength of the note, with the lowest line indicating the fundamental note while the parallel lines above represent the overtones that constitute the timbre for the voice. On closer listening, one can detect subtle shifts away from the pitch towards a more guttural enunciation at the end of each line; the vocal tone here assumes a sobbing quality, notably in the fading-out intensity of coloration at the ends of words like 'were' and 'burn'.

---

[7] Lori Burns theory on vocal presence and vocal authority is exemplary in understanding this phenomenon (Burns 2010).
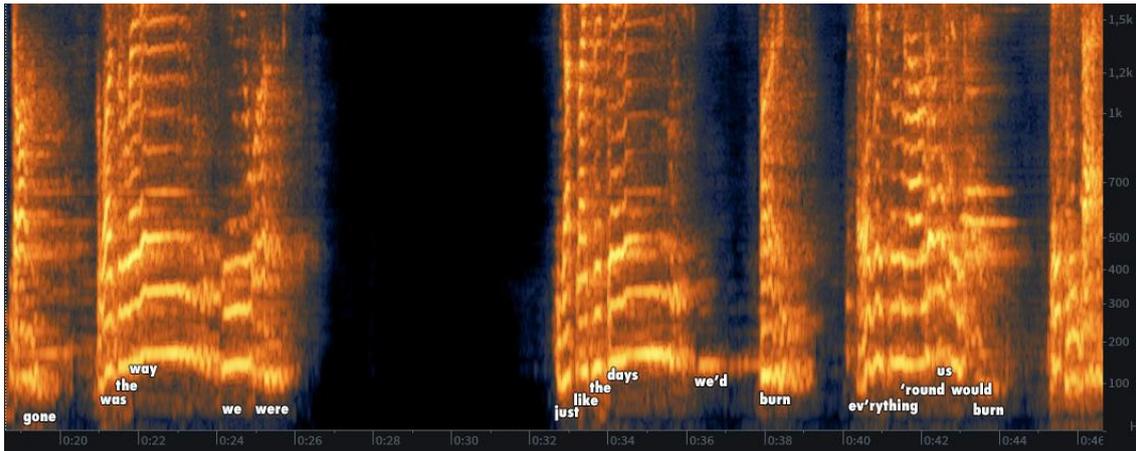
Figure 6: Spectrograph of 'Hope To Die' first verse, voice isolated

Spatiality in the recording and the voices position in the mix is shaped by a heavy use of reverb. Unlike the guitar and drums, the voice is not distanced from the front of the audio image by reverb. A subtle use of delay on the start of the reverb comes across first dry before being drowned in reverb. This is one of numerous instances where reverb takes over as the dominant sonic marker, leading to the impression that the "acoustical environment, in essence, becomes the sound source" (Moylan 2002, 266-7). Vocal presence is felt through the long pauses that not only create a vast sense of space in the song's mix, but also strengthen the relationship between the song's musical and lyrical structure. In the pre-chorus, the vocal track gradually increases in volume in anticipation of the chorus. During the second phrase, Peck goes up in pitch for the first time in the track on the phrase that begins with 'take me back'. At this moment, the pause following the word 'back' is elongated for almost a full measure. This arrival point, following a minute of melodic material on the first five notes of the A-major scale, involves an abrupt octave leap; the effect of this is to heighten the drama quality. Performatively, vocal techniques such as these and their audio engineering heighten the sense of a 'vulnerability-on-display', drawing attention to the role and arguably fragility of the singer.[8] As the vocal phrase extends on the lyrics, 'take me back, the word's I'd say, I had to whisper, because you liked it that way,' the earnestness in vocal expression is veritably hyperbolic.

This raises the matter of vocal strategies and the employment of stylistic codes within the track's overall compositional design. We have noted that Peck's

---

[8] See Hawkins' concept of vulnerability on display and masculinity (2009).

melodic lines are delivered in a quasi-operatic style through the use of deep vocal vibrato and a low bass register that is reminiscent of *Sprechgesang*. One might describe this as a form of *recitativo secco*, where the singer liberates rhythmic, melodic and harmonic structures by route of a melismatic approach. In turn this draws attention to melodic rhythm, which is enhanced by the slow tempo and relatively narrow pitch range that extracts connotations of this traditional style. An intertext, when it comes to tone, tempo, and timbre, is that of Johnny Cash, whose ability to delve into the low register while controlling long held notes with wide vibrato (a very difficult task) to accomplish the delivery of heartache, sorrow, and regret.[9] In Peck's vocal style there is also a nod to Depeche Mode's Martin Gore. His song, 'To Have and To Hold', involves a vocal delivery that comprises similar strained vocal folds. In the verses of 'Hope To Die', Orville Peck turns to timbral qualities in his vocal fold that emphasize exhaustion and heartbreak, where the voice almost cracks at the end of each phrase. Virtually evaporating, it leaves us hung and dry with little more than the faint sound of a scratchy throat. This is one of many examples where Peck vocal expression and positioning in the mix creates a vivid presence.

The structure and harmonic flavor in 'Hope to Die' are specific features of the recording. The main melody (Figure 2), first played by the electric guitar in the introduction, is repeated by Peck who repeats it in each of the verses. Commencing with a simple structure, it is elongated and languid: the first notes of each measure are sustained, with an emphasis on the first, fifth, second, then fourth degrees of the home key, A major. Upon Peck's entry, the harmonic accompaniment is taken by the electric guitar. In the second measure the mixolydian flavor is created by the minor dominant (Em), further reinforced by a resolution from v to ii (Em to Bm) in the third measure, before the chord progression, IV-V-I (D major-E major-A major) firmly establishes an A-major tonal center in the fourth measure.

The interplay between mixolydian and diatonic progressions for much of the track provides the song with its specific character.[10] To explain: while the song is in a major key, most of the chords are minor, creating a pull towards the mixolydian where the emphasis falls on the dominant minor and supertonic minor (v and ii). Driven by a lethargic tempo, the occurrence of chords, A-G#m(C#m)-

---

[9] See Askerøi 2017 for a detailed interpretation of Cash's voice in the track 'Hurt' (2002).
[10] For a discussion of modal and tonal ambiguity, see Hawkins 1992, 2002.

F#m-E suggest an Andalusian cadence,[11] commonly borrowed by blues and blues rock music, albeit in the major key. Arguably, while such harmonic elements are common to traditional blues, R&B, and country, there is a process of blurring on multi-levels that becomes a strategy of dramatization.
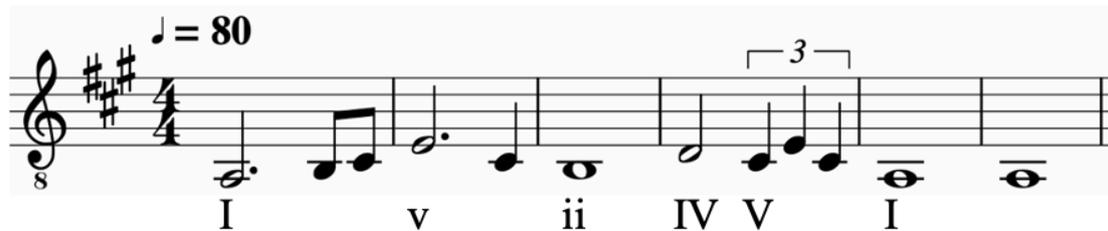

Figure 7: Primary melody of 'Hope to Die' with harmonic labels

In terms of the song's stylistic codes, there is an attempt to break away from the conventional structural norms found in country and folk music. This is discernible in the pre-chorus (0:43-1:16 and 2:05-2:39) that begins with the standard chord progression, I-iii-vi-V. On repetition, however, it is cut short by one measure with a long IV chord that extends over two bars, followed by the dominant leading into the chorus. The effect of this break is dramatic, further heightened by the slowness in tempo, with the duration of each section fading out before the end. Unexpected breaks in the song's formal structural are eased by predictable chord progressions, another tactic of interest or surprise. The four measures following this break involve exaggerated pauses, filled in only by the voice and a single guitar strum that offers slight harmonic support. In Figure 3, the song's form is mapped on to aspects of harmonic structuring to illustrate the subtleties of compositional design.

During the final lap of the song (from measures 62-71), yet another departure from traditional formal devices occurs now in the form of an instrumental bridge (ca. 28 seconds), preceded by an unexpected break as the chorus is cut short by one measure and paused. In the music video, such gestures are used to dramatic effect, as we will turn to later in the essay. While the harmonic rhythm at this point remains constant, the guitar reiterates the chords from the verse, but this time

---

[11] The Andalusian cadence is a descending tetrachord with its origins in the Flamenco tradition, either as iv-III-II-I (phrygian) or i-VII-VI-V (minor), typically signaling the end of a long section or a piece of music in an ostinato form. For more on the cadence and Flamenco guitar playing specifically, see the thorough guide "Music Theory for Flamenco" by Chuck Keyser (1998). Also for a discussion on the appropriation of Flamenco's sounds in popular music generally, see Folch 2013.

omitting any modal flavor by use of a major dominant. Rhythmically, though, both the guitar and percussion convert into double-time, with a quasi-Latin strummed dance groove alternating between the guitar and dry rim shots.

Tempo: 80bpm – Key: A

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Introduction (m. 1-6)						Verse 1 (m. 7-14)								Pre-chorus 1 (m. 15-25)										
a (4)						a (4)				a (4)				b (4)				b' (3)			c (4)			
I	v	ii	IV	V	I	I	v	ii	IV V	I	v	ii	IV V	I	iii	vi	V	I	iii	vi	IV		V	

26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Chorus 1 (m. 26-33)								Verse 2 (m. 34-41)								Pre-chorus 2 (m. 42-52)								
d (4)				d' (4)				a (4)				a (4)				b (4)				b' (3)			c (4)	
I	vii	vi	V	I	iii/G#	vi	IV	I	v	ii	IV V	I	v	ii	IV V	I	iii	vi	V	I	iii	vi	IV	

51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
		Chorus 2 (m. 53-60)								Pause ~7 sec	"Ballet" Dance break, x2 time feel (m. 62-69)								Shot!	Chorus 3 (m. 71-78)				
		d (4)				d' (4)					a"				a"					d (4)				d' (4)
V		I	vii	vi	V	I	iii/G#	vi	IV		I	V	ii	IV V	I	V	ii	IV V	— *	I	vii	vi	V	I

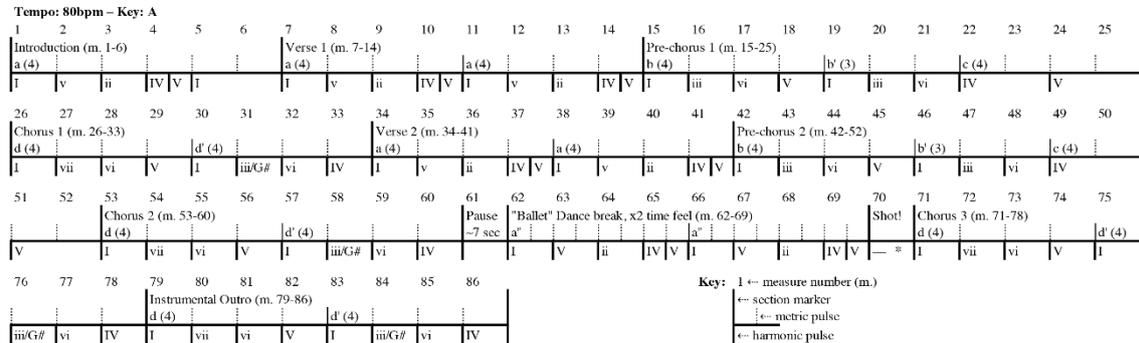| 76	77	78	79	80	81	82	83	84	85	86		
			Instrumental Outro (m. 79-86)								**Key:** 1 ← measure number (m.)	
			d (4)				d' (4)				← section marker	
											← metric pulse	
iii/G#	vi	IV	I	vii	vi	V	I	iii/G#	vi	IV	← harmonic pulse	

Figure 8: Formal Structure of 'Hope To Die'

As the section draws to a close the dominant chord rejects resolution, cut short by a pause that is filled by loud drum parts leading into the final chorus, modulating from A major to C-major. Melodramatic and arguably clichéd, this modulation is a stylistic code commonly found in the songs of female pop singers, such as like Whitney Houston, Madonna, and Celine Dion, for whom such theatrical harmonic devices are standard practice.[12] While the opening part of the chorus melody throughout the song is a reference to the chorus of the song 'I Will Always Love You' by Dolly Parton, the elevated modulation, pause and drum lead-in is in no uncertain terms a more striking reference to Whitney Houston's version of the same song, with a similar compositional device announcing the final chorus.[13]

The slow tempo of 'Hope to Die' affects the overall timbral quality of the recording, which is critical to our music analysis. Excruciatingly slow, the pace poses inevitable vocal challenges. Not unlike Steven Morrissey, who is known for slow, mournful songs, Peck's approach to melodic tempo raises a host of issues that relate to control and regulation. His 'pacing' is characterized by lengthy pauses, 'erotic gaps', that result in astute dramatic intensity. The entire sound production is greatly sculpted by effects, such as reverb; in a sense, this has a

---

[12] Modulations of this ilk have often been disparaged and dismissed as kitsch, trivial, and overkill. See the writings on 'bad music' by writers Robert Walser, Carl Wilson, Simon Frith, Dai Griffiths, Stan Hawkins, Washburne and Derno.

[13] Orville Peck has acknowledged the profound influence of Whitney Houston on his musical style and identity.

manipulative effect by subverting a range of norms that pertain to country music. Through the elements of the mix, we gain an insight into how Orville Peck masters the technologies of music production.

Figure 9: Transcription of the bridge of 'Hope To Die', mm. 62-69

The production certainly contributes to the narrative effect. Comprising relatively sparse instrumental material, the arrangement consists of two electric guitars, kick and snare drums, bass guitar, and voice. In the verse sections there are two main guitars: a rhythm guitar outlines the harmonic rhythm, while the other guitar extracts the melody at the end of each phrase. In the mix the guitars are awash with reverb, giving the sensation of a large and empty concert-hall space. The rhythm guitar line is double tracked in stereo, while the melodic guitar is mono, centered within the mix alongside the lead vocal, processed through a stereo reverb that increases its apparent size within the mix.

So far, we have suggested the regulation of vocal and instrumental timbre is a central stylistic device in 'Hope to Die'. In a sense the implementation of tempo and timbre stylistically imitates operatic style. Similarly, the ways in which the electric guitars sound is perceived is contingent on the tempo. The rhythm guitar, emphasizing each chord change with slow strumming, is an electric guitar recorded

through a Fender Deluxe Reverb Amplifier, or a similar amp with reverb and chorus settings. While the sound is characteristic of the classic pedal steel guitar tone found in much country music, in this case it is not performed by a pedal steel. In sounding out the chords, the guitar, as with the voice, leaves a large amount of space filled by the sounds not only by reverb but in the case of the guitar a subtle chorus effect, likely of the amplifier, which would be otherwise almost imperceptible but for the sparse nature of the arrangement. In staking out the melodic material in tandem with the vocal part, Notably, the lead melodic guitar is foregrounded in the mix, panned in mono but with a large stereo reverb, occasionally competing with the voice for space in the mix. The guitar timbre at this point is very 'clean', with a heavy 'twang' in its equalization—characteristic of guitar in country music to be sure. The heaviness of the pick by the performer, combined with the often-audible portamento between pitches in the melody, and the subtle boost in the lower mid-range combine to give a sound that is reminiscent of classic americana, such as that in Johnny Cash's 'I Walk The Line' or The Highwaymen's self-titled single 'Highwayman'.

Finally, mention should be made of the drum parts: the rhythm established at the beginning persists throughout the entire track, with little variation except in the short bridge toward the end. The effect of reverb on the drum sound produces a 'long tail', made all the more audible by the sparse nature of the beat. The beat itself is compatible with the overall aesthetic of the track, and the slow, repeating dotted eight-note pattern of the kick is reminiscent of a heartbeat, which can be interpreted as somber and relaxed. Its dramatic function is important, and the slow tempo is enhanced by a sense of exasperation in the release of the thick-toned snare drum with its massive reverb that fills the audible empty space. Especially noteworthy is the reverb on the snare drum, which is magnified in the chorus sections, with a slight rise in pitch. Timbrally, the use of a different drum in these sections heightens both the energy and drama of the chorus.

Sonic Imagery

We now turn to a consideration of the relationships between sound and visual staging in the video of 'Hope to Die',[14] our method being to equate elements of the

---

[14] Official video: https://www.youtube.com/watch?v=60MHmrtEuRY

sound recording with the visual dramaturgy.[15] As such, we examine the storyline, recording elements, and compositional elements to bring to the surface the aesthetics of Peck's performance. To start with, we suggest that the visualization of human performance in recorded format is charged with impressions of the audio track. Lori Burns and Stan Hawkins, in the introduction to the *Bloomsbury Handbook of Popular Music Video Analysis*, state: "The sight of the performing body invites intensified reflexivity on the part of the viewer, characterized by embodiment, simulation, cognition, and agency" (Burns & Hawkins 2017, 3). As we have hitherto pointed out, Peck's vocal performance in 'Hope to Die' is awash in timbral detail, resonating in a manner that impacts the visual effects through a catalogue of movements. Every section of the song possesses its own contours, which are set off by the intensity of every image frame, whereby the elements of audiovisuality have a powerful bearing on one's perceptions of sound. By referring to elements, such as color, lighting, choreography, fashion, props, and scenery, we have extracted four main moments in the video that aid our analytic observations:

Introduction (0:00-0:55): Starting with the two men silhouetted against a white backdrop, we gain the first sight of Peck who is having his mask adjusted in preparation for the performance. In terms of subject-positioning, there is a reference to Jim French's iconic portrait of two cowboys in the same stance, naked from the waist down, from 1969, which would be later popularized by the Sex Pistols' bassist Sid Vicious.[16] The opening shot is homoerotic, reinforced just 30 seconds later by a low shot angle from the floor in between the bare muscular legs of another cowboy. The second cowboy, credited as Adrian Nallo, has a candid expression, and is dressed in loose-fitting yet suggestive clothes that obfuscate to some degree his physique. The depiction of masculinity here queers the hetero norm, a common trope in Peck's other videos, such as 'Queen of The Rodeo' and 'Dead of Night'. The first frame comprises twenty seconds with the instrumental Johnny Cash-like guitar melody, heavily reverbed, creating a high sense of expectation. Peck's costume is tight fitting, resembling that of a Spanish matador (apart from the cowboy hat and mask). The first scenes with him function as a short prelude to the commencement of his dance, whose choreography draws on the

---

[15] To date much has been written on audiovisuality in pop music where emphasis is placed on the creative enterprise of performance. See Auslander, 2008, 2009, 2021; Burns & Hawkins, 2019; Burns & Lafrance, 2017; Burns & Watson, 2010; Burns & Woods, 2018, 2019; Hawkins, 2004, 2009, 2016; Korsgaard, 2013, 2019a, 2019b; LaFrance, 2013; Lafrance, Burns, & Woods, 2017; Vernallis, 2008, 2019.

[16] http://www.paulgormanis.com/?p=2603

traditional Paso Doble, with an emphasis primarily on his hand and arm movements; they are circular, extending above the head and forming a rounded contour as he starts singing. With little warning the camera angle suddenly changes (0:40) panning to floor level as we view Peck through the legs of what first seems to be another male about to move into combat mode. The homoeroticism of this shot is short-lived via a medium zoom through the legs on to Peck, as the lens angle is changed and close-up. At this point the lighting changes as Peck is silhouetted against a black backdrop with one spotlight pointed over him. Shots of him singing with a stand microphone are then juxtaposed with ponies (in close up), as he completes the phrase, 'take me back to the time you were mine' (0:45-0:55). On the words 'you are mine' there is a costume change with new head-dress (with a brim shaped to look like bull horns protruding from his hat) as he is filmed holding two ponies by their reigns.

Chorus 1 (1:18-1.42): Peck faces the camera through the arch of open legs, with his shirt unbuttoned, exposing a muscular torso. The blatant reference to gay cowboy pornography is indisputable, a full front confrontation of the male, cisgender gaze. This shot (1:18-1:27) is in color, with Peck sporting a long, blue fringe mask that matches the denim shirt and trousers that have a double clasp belt, with silver sheriff badge. Notably, the rest of the chorus is filmed in monochrome, its visual aesthetics defined starkly by silhouette shots against a white screen, including the protagonist on his own, then with the two brown ponies, and, finally, him standing back-to-back with the other cowboy. The visual hue alternates between shades of color, matching the subtleties of sound production. Overall, an increase in rhythmic visual movement characterizes the choreography, particularly in the hand, arm, and torso gestures, with the contrasting textures and timbres in the mix. On close inspection, Peck's repertoire of movements suggests a stance of defiance through an arching of his back and head backwards and a raising of his right arm slowly upwards. His choreography involves agile yet slow motions of stretching the upper torso that are reminiscent of flamenco; the dramatization of this spectacle is intensified by the drawn-out melodic lines (vocally and instrumentally) and heavily reverbed production. In many ways the musical material complements the feline qualities of hyper-masculinized imagery. Homoerotic codes reference the legendary art of Tom of Finland, whose trademark was built upon exaggerated traits of tight and partially removed clothing, sexual allure, emphases on muscles and genitalia, and tough and tender depictions of S&M. Poignantly, this chorus dissolves with Peck dropping his head, as he belts

out the mournful phrase, 'cross my heart, now I hope to die' (1:38-1:42), a hard accent placed on the word 'die'.

Instrumental Bridge (3:09-3:36): The vocals drop out as the guitar and percussion take over. Peck's visual agency is at its most powerful here as a solo dance ensues in a dimly lit barn with the light streaming in from a skylight in the ceiling. In this scene the details of his face are obscured. Something chilling in his aura draws on the score of Brokeback Mountain, evident in the instrumentation and epic cinematic feel. Another costume change occurs with Peck in a white short-sleeve t-shirt, an extra-long fringe mask and tight-fitting trousers with 'chaps' – coverings for the protecting the legs made in leather-like material named after the Spanish *chaparral* for brush, thorny, and thick, designed to protect the legs when riding horseback and bull-riding in rodeo man culture. Significantly, this accessory has played a major part in the assimilation of cowboy culture into the American West. For the first time, the camera focuses on Peck's lower torso and legs, as he executes his own version of a barn dance. The sequence starts with him stomping his feet like hooves in the straw and dust, revving up for a genre-bending repertoire of moves that stylistically challenge anything associated with cowboy traditions of moving the body to music: we witness a crossover of ballet, line dance, disco to MJ pop. Starting with the sound of his stomp on a wooden floor, the rhythmic part gives way to a rapid fire of *palillos* or castanets, which belong to the clapper group of percussion idiophone. This figure, which accompanies the main guitar melody (see Fig. 4), is the most frenetic in the entire track matching the dance movements that showcase the virtuosic skills of Peck. While some of his foot movements might well belong to dance routines associated with cowboys, such as the hands on the hips with step movements, the pirouettes derived from classical ballet certainly do not, where the dancer rotates on one leg with the other off the ground. In order to maneuver this exceptionally difficult move, dancers need to be graceful, flexible, and sturdy, as it involves a complete turn of the body on one foot (*en pointe*). Peck delivers three pirouettes (3:17, 3:29, and 3:36), the final one culminating with his arms reaching a wide-open gesture through circular motion. The effect of such visual spectacle is to prepare us for the climactic point (see Fig 5), which starts with a long pause and then an elevated modulation.

Figure 10: Mid-pirouettes during the barn ballet dance

Final Chorus and Outro (3:36 to end): A shot of Peck's hand imitating a revolver shot triggers the scene change to the artist on stage, with upper torso and head on display. This signals the climactic point of the song, following the elevated modulation, with a rise in intensity as Peck uses wide open arm gestures that circle above his shoulders and head. The lighting creates a semi-silhouette effect, with the red mask and inner lapel contrasting with his black costume. As he sings, 'cross my heart now to die' (3:59-4:03), his costumes change twice from the blue denim and blue mask to a rhinestone leather jacket. The scene with blue denim refers back to the first moments in the video, where he is framed by the leg close-up of his partner. As he beats his breast in one of the shots (4:06-4:11), his hands are filled with a white milky jizz-like substance that splatters over his black shirt. The sexual reference is overtly tongue-in-cheek and rife with humorous intent and sex innuendo. Then yielding to an earlier image of him with the two brown ponies, the next scene includes shots of him standing back-to-back with Adrian Nallo. In the final scene he is seen collapsed, lifeless, on the floor with stars dangling down in front of the black curtain drop; his name PECK is positioned at center top. The audiovisual tempo of this final chorus is painfully slow— the long-sustained notes brim with vibrato, lush and open reverberation, slow and deliberate movements, dramatic and long-held poses—all of which support the visual representation of a tormented soul with a broken heart. Highly dramatic, this scene (4:23-4:38) culminates with the main guitar melody and then a few seconds of silence, giving the viewer pause for thought. The imagery is rendered all the more powerful by the resonance of the highly reverbed final note (pitch C) and its slow diminuendo into nothing but silence. Startling in its degree of intimacy the final shot

155

exaggerates Peck's vulnerability in the form of the crushed cowboy strikes back at the restrictive gender roles we are familiar with in popular culture. A complicated assimilation of the queer cowboy into country culture is epitomized in these final seconds of sheer abandonment. Few pop videos entertain such closure.

In these four moments of close readings, we are mindful of the multimodal aspects of expression within close readings (Burns 2018, pp. 95-97). We concur with Burns' definition of multimodality as "the artistic integration of multiple semiotic modes within one media text" distinct from the concept of multimedia wherein multiple texts are present in one setting (Burns 2016, pp. 96-97). Applied to our reading of 'Hope to Die', such semiotic heterogeneity becomes apparent—not only in the modes of sound and image, but also in text, dance, fashion, and vogueing.[17]

Art of Masking: aesthetics and production

Contemplating visual features is an all-defining element of gaining meaning from the pop score. Connecting recording elements in the track's structure to the imagery of Peck's performance in the video is relevant to the listening process in much pop music. At the core of our task is the matter of understanding the intentions of the artist. In the track it is the multidimensionality of the audiovisual production that prompts observations and then evaluations. We concur with Moylan that "a multidimensional, multi-domain texture exists at the highest dimension of the recorded song; this is the overall sound quality of the track" (2020, p. 197). By engaging with textural domains in 'Hope to Die', one of our aims is to unravel the aesthetics of production that are attributable to an array of features: sonic invention, production techniques, compositional design, lyrical treatment, and performance strategies, which we now turn to.

Our findings indicate that vocal presence and attributes of singing, stylistically and technically, shape the mood of the track, with Peck turning to a range of techniques to convey a high level of pathos. His layering of stylistic references is integral to the sonic inventions in compositional design. In this way, the dramatic quality of the song is harnessed by the arrangement, where varying degrees of spatiality heighten the levels of interest on the part of the listener. As we have

---

[17] Similarly, Mathias Korsgaard has insisted that music video is "defined by its very heterogeneity, its wide range of audiovisual expressions" (2017, p. 30).

indicated earlier, this impacts on harmonic and melodic structuring, and the subtleties of coloration that are complemented by rhythmic details, such as the guitar and percussion parts during the final measures of the track (Figure 4) in the form of the quasi-Latin groove. The creative regulation of the sound stage, as we have argued, draws attention to the vocal presence in relation to the surrounding textures and timbres as stage undergoes continual transformation. Accordingly, the recording and sound production shapes the narrative. The vitality of Peck's vocal part in terms of detailed sonic processing provides an illusory sense of reality, a sense of masking, that is conditional on the content and character of what Moylan terms the 'holistic environment', which "establishes an expanse that complements the expression of the track" (2020, p. 478). Integral to this environment is the sonic invention of spatiality that is moderated creatively to form a rich aesthetic backdrop.

Figure 11: The art of masking as vogue

Arguably, the deceptively simple arrangement strikes an intricate balance with the technologically sophisticated performance. For instance, the reverb and delay effects not only serve as placeholder for the long gaps between rhythmic and harmonic events in the score, but also amplify the performative drama and vogue-like posturing in both the music recording and music video. The theatrical use of reverb is most salient at the moment of modulation going into the last verse (3:35-3:45), when the progression is cut off by a measure, creating a silence filled only by reverb before the drums pound into the last section. Of significance is also the close-up sense of vocal staging and signal processing, which establishes a high degree of intimacy in stark contrast to the epic-like spatiality afforded by reverb.

Clearly audible in the vocal track especially at the ends of phrases in the verses is the tapering of long-sustained tones that give way to 'scratching fry sounds' that are emulated through vibrato. This raises notions of masking, which in music production can refer to the effect where "a sound (or portion of a sound) is not perceived because of the qualities of another sound" (Moylan 2002, p. 32). Similarities in pitch range between the guitars and the voice in such moments of 'vocal fry' create a sense of drama in the voice that is heightened as it is slowly masked by the sounds which come to dominate. Such sonic effects are a result of the combination of close-microphone techniques, processing through vocal compression, and clever mixing. As such, highly produced vocal sounds evoke the emotional sense of anguish and vulnerability and reinforce visual interpretations of masking.

In keeping with many pop music recordings and videos, 'Hope to Die' is replete with signifiers that entice the listener/viewer. In our hermeneutic approach we recognize that the audiovisual codes of the pop score coalesce at the moment of reception to create a propensity for immersion. In particular, Peck's vocal styling is a primary teaser: his low register, combined with deep vibrato and close-microphone techniques, often blurs the line between singing, speaking, and crying creates intertextual reference not only to the vocal stylings of mid-century country and folk singers, such as Johnny Cash, Dolly Parton, and Willie Nelson, but also as we have shown to the drama of opera, Sprechgesang, and recitativo secco. In terms of the track's compositional design, the lethargic rhythmic and harmonic pace taunts a mixolydian major/minor uncertainty through the use of the blues Andalusian cadence. The standard verse-chorus structure is also cleverly broken up at times creating dramatic and unexpected shifts between sections and making space for pregnant pauses. And, as we have demonstrated, the slow tempo highlights timbre, spatial sound, and balance as the most important compositional elements in the track.

As a rule, vocal style is an outgrowth of recording techniques and decision-making. In 'Hope to Die' this is evident in an intensity of expression that feels constructed in a way one never encounters in concert experiences. One might argue that a sense of over-production makes the song knowingly artificial and non-real. Yet, as Virgil Moorefield insists, in pop recordings realism is not the point: "What matters is the sonic experience the record offers, on its own terms, as *sound*" (Moorefield, 2005, 55 – author's emphasis). The recording and its concept draw attention to not only the meaning of what is being sung, but also to the aesthetic

effects of the sound recording. In the song under analysis, poignancy is rendered mainly by the chorus hook, where the final words 'hope to die' are delivered with soft dynamics in the lowest register. Notably in the second chorus the word 'die' is omitted as Peck delivers 'now I hope to…' (3:02), which is followed by a drop out of all musical material with just the sound of footsteps on the barn floor (3:03-3:10), which then leads to the instrumental passage with tap-dancing sounds complementing the guitars. Following this passage, the climactic moment is unleased, 'I'm still undone', with the final chorus where the phrase, 'now I hope to die' (4:02-4:14) is repeated twice before the track ends. Quivering on the word 'die', the vocal part gives way to the guitar which plays the melodic hook one last time. The crooning quality of Peck's voice defines his persona, and this is conveyed by degrees of timbral coloration that fluctuate between a sense of tranquility and urgency that impact on the narrative of the song. Peck's highly produced voice defines what Moylan has described the 'performance intensity' (2020, p. 451), where the verbal space of the mix is contingent on the timbral and textural employment of distance positions. All this occurs within a holistic environment, where the technicalities of stereo positioning, effects, compression, equalization, and microphone spell out the attributes of the pop recording.

Conclusion

As we have attempted to demonstrate, Peck's performance in 'Hope to Die' furnishes a narrative that brings things to life in an extraordinary fashion. We have found that his distinctly queer sensibility characterizes a mode of articulation that is highly contingent on the multiple details of the recording process. In addition, the video contains a wealth of elements that result in mediating the antics of the gay cowboy, which we interpret as a liberation from stereotype strictures of representation. Any notions of utopia are short-lived, however, by the protagonist's hope to die. Nestled within the spatial environment of the recording is the staging of a queer performativity with a tinge of humorous intent. Peck sings, 'But I, I still try, cross my heart, now I hope to die' with an air of melodrama, a plea to inspire a social and political sense of becoming; as such, he taunts difference as a facet for openness and promise.

These sentiments are delivered passionately in 'Hope to Die'. In the end, Peck's performativity calls into question an alternative political impasse for understanding the past by putting into play the ideology of gender difference, all of which is achieved by an audio effect of hollowness that is down to the spatial

expanse of the mix and production. Alas, freedom lies in the hope to die with a gravitational pull away from utopian longing. Peck's strategy of cruising a narrative in a controlled audio space might not only be a response to repressive heteronormativity, but also a bolstering of pop ephemera in the playfulness of the queer cowboy!

References

Askerøi, E. (2017). Spectres of Masculinity: Markers of Vulnerability and Nostalgia in Johnny Cash. In S. Hawkins (Ed.), The Routledge Research Companion to Popular Music and Gender (pp. 63-76). New York: Routledge.

Auslander, P. (2008). Liveness: Performance in a Mediatized Culture (2nd ed.). New York: Routledge.

Auslander, P. (2009). Musical Persona: The Physical Performance of Popular Music. In D. B. Scott (Ed.), The Ashgate Research Companion to Popular Musicology (pp. 303-315). Surrey, UK: Ashgate.

Auslander, P. (2021). In Concert: Performing Musical Persona. Ann Arbor, MI: University of Michigan Press.

Burns, L. (2010). Vocal Authority and Listener Engagement: Musical and Narrative Expressive Strategies in the Songs of Female Pop-Rock Artists, 1993–95. In M. Spicer & J. Covach (Eds.), Sounding Out Pop: Analytical Essays in Popular Music (pp. 154-192). Ann Arbor, MI: University of Michigan Press.

Burns, L. (2016). The Concept Album as Visual-Sonic-Textual Spectacle: The Transmedial Storyworld of Coldplay's Mylo Xyloto. IASPM Journal, 6(2), 91-116.

Burns, L. (2018). Interpreting Transmedia and Multimodal Narratives: Steven Wilson's "The Raven That Refused to Sing". In C. Scotto, K. Smith, & J. Brackett (Eds.), The Routledge Companion to Popular Music Analysis: Expanding Approaches (pp. 95-113). New York: Routledge.

Burns, L., & Hawkins, S. (2019). Introduction. In L. Burns & S. Hawkins (Eds.), The Bloomsbury Handbook of Popular Music Video Analysis (pp. 1-9). New York: Bloomsbury.

Burns, L., & Lafrance, M. (2017). Gender, Sexuality, and the Politics of Looking in Beyoncé's 'Video Phone' (Featuring Lady Gaga). In S. Hawkins (Ed.), The Routledge Research Companion to Popular Music and Gender (pp. 102-116). New York: Routledge.

Burns, L., & Watson, J. (2010). Subjective Perspectives through Word, Image and Sound: Temporality, narrative agency and embodiment in the Dixie Chicks' video 'Top of the World. Music, Sound, and the Moving Image, 4(1), 3-37.

Burns, L., & Woods, A. (2018). Rap Gods and Monsters: Words, Music, and Images in the Hip-Hop Intertexts of Eminem, Jay-Z, and Kanye West. In L. Burns & S. Lacasse (Eds.), The Pop Palimpsest: Intertextuality in Recorded Popular Music (pp. 215-251). Ann Arbor, MI: University of Michigan Press.

Burns, L., & Woods, A. (2019). Humor in the "Booty Video": Female Artists Talk Back Through the Hip-Hop Intertext. In T. M. Kitts & N. Baxter-Moore (Eds.), The Routledge Companion to Popular Music and Humor. New York: Routledge.

Folch, E. (2013). At the Crossroads of Flamenco, New Flamenco and Spanish Pop: The Case of Rumba. In S. Martinez & H. Fouce (Eds.), Made in Spain: Studies in Popular Music (pp. 33-43). New York: Routledge.

Hawkins, S. (1992). Prince: harmonic analysis of 'Anna Stesia'. Popular Music, 11(3), 325-335.

Hawkins, S. (2002). Settling the pop score: Pop texts and identity politics. Burlington, VT: Ashgate.

Hawkins, S. (2004). On performativity and production in Madonna's 'Music'. In S. Whitely, A. Bennett, & S. Hawkins (Eds.), Music, Space and Place: Popular Music and Cultural Identity. Surrey, UK: Ashgate.

Hawkins, S. (2009). The British Pop Dandy: masculinity, popular music and culture. New York: Routledge.

Hawkins, S. (2016). Queerness in Pop Music: Aesthetics, Gender Norms, and Temporality. New York: Routledge.

Keyser, C. (1998). Music Theory for Flamenco [webpage]. https://www.flamencochuck.com/files/Music%20Theory/Theory.pdf. Accessed 28 June 2021.

Korsgaard, M. B. (2013). Music Video Transformed. In J. Richardson, C. Gorbman, & C. Vernallis (Eds.), The Oxford Handbook of New Audiovisual Aesthetics. Oxford: Oxford University Press.

Korsgaard, M. B. (2019a). Changing Dynamics and Diversity in Music Video Production and Distribution. In L. Burns & S. Hawkins (Eds.), The Bloomsbury Handbook of Popular Music Video Analysis (pp. 13-26). New York: Bloomsbury.

Korsgaard, M. B. (2019b). SOPHIE's 'Faceshopping' as (Anti-)Lyric Video. Music, Sound, and the Moving Image, 13(2), 209-230.

LaFrance, M. (2013). Celebrity, Spectacle, and Surveillance: Understanding Lady Gaga's 'Paparazzi' and 'Telephone' through Music, Image, and Movement In M. Iddon & M. L. Marshall (Eds.), Lady Gaga and Popular Music. New York: Routledge.

Lafrance, M., Burns, L., & Woods, A. (2017). Doing Hip-Hop Masculinity Differently: Exploring Kanye West's 808s & Heartbreak through Word, Sound, and Image. In S. Hawkins (Ed.), The Routledge Research Companion to Popular Music and Gender (pp. 285-299). New York: Routledge.

Moorefield, V. (2005). The Producer as Composer: Shaping the Sounds of Popular Music. Cambridge, MA: MIT Press

Moylan, W. (2002). The Art of Recording: Understanding and Crafting the Mix. New York: Focal Press.

Moylan, W. (2020). Recording Analysis: How the Record Shapes the Song. In. New York: Routledge.

Vernallis, C. (2008). Music video, songs, sound: experience, technique and emotion in Eternal Sunshine of the Spotless Mind. In Screen (Vol. 49, pp. 277-297).

Vernallis, C. (2019). 12 Writing about music video. In Writing About Screen Media (pp. 12).