

Sound Events, Spatiality and Diegesis – The Creation of Sonic Narratives in Music Productions

Introduction

“You Think I Ain’t Worth a Dollar, But I Feel Like a Millionaire” is the opening track on Queens of the Stone Age’s third studio album *Songs for the Deaf*.¹ The track begins with the sound of clanking keys, a car engine that starts and the alerting beeps warning the driver that doors are still open. A car radio is switched on and the car door is slammed bringing the beep sound to a halt. The radio is tuned and eventually settles on a station where a male voice announces: “Los Angeles, Clone Radio, we play the songs that sounds more like everyone else, than anyone else.” The song following the presentation opens with a steady drumbeat. Seconds after, a distorted guitar riff joins in. What we are witnessing is a rock song, transmitted to us through a car radio; an effect that emerges mainly as a result of low cut filtering in the 200 Hz region and a high cut around 7kHz that creates the impression of a small lo-fi stereo-system. The rock song is there for the listener as something played back into the virtual car setting, and thus present only in a double-mediated sense.

There is a sudden change in the scene the moment vocals² and bass guitar enter at 1’01 where the track abruptly increases in loudness and the frequency band increases to full spectrum. The perceptual effect of this change can be described in at least two ways. Either the radio track “pops out” of the car speakers, entering and now residing in the virtual listening space (the car setting) that was previously established; or the track “drags the listener in.” In both cases, the changing point of audition (POA) is not the result of the track moving closer to the listener across the virtual space of the recording. There is a change in the ontological³ level of the track and in this way the

- 1 Queens of the Stone Age. “You Think I Ain’t Worth A Dollar But I Feel Like A Millionaire,” *Songs for the Deaf* (Interscope, 2002). The album is produced by Josh Homme (front singer) with record engineers Adam Kasper and Eric Valentine.
- 2 Lead vocals on this track is Nick Oliveri and not Queen of the Stone Age’s usual front singer Josh Homme.
- 3 Building on the viewpoint taken by Merleau-Ponty that the body-subject is intimately tied to lived space and, consequently, that the phenomenal world equals the real world, my use of the term ontology refers to the categories (or hierarchies) of spatial being in the recording (e.g. that is how, and at what ‘level’, spatial entities in the recording can be said to exist) that is formed in perception – a

track transcends the virtual borders (in time and space) of the radio's virtual space, thus creating a dramatic increase of presence for the listener.

The example above serves to illustrate two points that will be dealt with in this article: 1) the potential for sound engineers to radically effect the listener's perceived point of audition, and create different forms of tensions and contrast effects between these different perspectives; and 2) the ontology of recorded sound (i.e. the time-space continuum of the perceived auditory events) is only fully realized if we consider how sound events coexist within the imagined performance space of recordings. The purpose of this article is to discuss the relation between sound qualities emerging from studio-based editing and the perceived diegetic framing in a number of popular music productions. I revisit the concept of diegesis in this context to explore its applicability in the analysis of recorded music and as a tool to conceive of the perceptual effect emerging from recording and editing practices.

For several decades, film sound theorists have used the concept of diegesis to explore the relation between sound (e.g., background music, dialogue, voice over) and image in film.⁴ More recently Anahid Kassabian has argued that the diegesis of the film is not necessarily determined *prior* to the music, but music itself can construct a diegetic framing depending on the degree of attention given to music in the particular scene.⁵ Also a number of studies have redefined the concept to explore the function of sounds in computer games.⁶ The reality of movies, like the reality of recorded performances, never existed as a real event. Both movies and music recordings consist of clips and layers of recorded material spliced together to form, in our experience, a coherent event. Nicholas Cook notes this comparison between movies and music recordings writing that:

(...) the concept of diegesis applies just as well to sound recordings [as to movies], and demonstrates the sense in which the relationship between the recording and

phenomenological ontology. Also – and this is in line with Brian Massumi's thinking – I think of reality as a part of virtual spatiality (e.g., something sensed as real through the virtual) rather than contrasted to it. See Maurice Merleau-Ponty, *Phenomenology of Perception* (London: Routledge and Kegan Paul, 1945/2002); M. C. Dillon, *Merleau-Ponty's Ontology* (Evanston, Illinois: Northwestern University Press, 1988); Brian Massumi, "Envisioning the Virtual," in *The Oxford Handbook of Virtuality* ed. Mark Grimshaw (Oxford and New York: Oxford University Press, 2014).

- 4 See David Bordwell, *Narration in the Fiction Film* (London: Routledge, 1985); Michel Chion, *Audio-Vision: Sound on Screen* (New York: Columbia U.P., 1994); Claudia Gorbman, *Unheard Melodies* (Bloomington and Indianapolis: Indiana University Press, 1987); Étienne Souriau, *L'Univers filmique* (Paris: Flammarion, 1953).
- 5 Anahid Kassabian, *Hearing Film: Tracking Identifications in Contemporary Hollywood Film Music* (New York: Routledge, 2001); Anahid Kassabian, "The End of Diegesis as we Know It?" In *The Oxford Handbook of New Audiovisual Aesthetics*, eds. John Richardson, Claudia Gorbman and Carol Vernallis (Oxford and New York: Oxford University Press, 2013).
- 6 Karen Collins, *Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design* (Cambridge: MIT Press, 2008); Mark Grimshaw, *The Acoustic Ecology of the First-Person Shooter: The Player Experience of Sound in the First-Person Shooter Computer Game* (VDM Verlag, 2008); Kristine Jørgensen, "On Transdiegetic Sounds in Computer Games," *Northern Lights* 5 (2007), 105-117; Ulf Wilhelmsson, *Enacting the Point of Being. Computer Games, Interaction and Film Theory* (PhD Thesis. University of Copenhagen, 2001).

the experience is a semiotic one (...) it is through the act of listening which the recording prompts that we understand it as a trace of an event.⁷

Also Simon Trezise argues that the concept is useful to analyze the purely auditory domain⁸ but still there is no in-depth theorization of how the diegesis of music recordings is formed and exploited in the recording studio.

In the first part of this article I discuss diegesis in relation to the production and perception of recorded popular music. Most producers seem to aim for mixes that produce the illusion of a single diegetic frame containing all sound events. For this reason, the best way to recognize the value of my argument is by exploring some of the exceptions – that is, tracks where the diegetic relationship between sound events is challenged or simply broken. In order to discuss the different ontological relationship between sound events within recordings I use Gérard Genette's classification of narrative events: the diegetic, meta-diegetic and extra-diegetic level,⁹ later used by Claudia Gorbman in her classification of film sounds.¹⁰ Although there are many similarities between how we make sense of audio-visual media and audio media, the notion of diegesis is not unproblematic in the context of popular music productions. For this reason, this article proposes a systematic theoretical framework to discuss the concept in relation to recorded music – what I shall here call a *phonographic diegesis*. The second part of the article addresses a number of musical examples to exemplify different modes of layering sound events.

Presenting the Story

In the third book of *The Republic* Plato contrasts narration (diegesis) with imitation (mimesis) as two ways of presenting literary works. In narration, the story world is presented through the narrator telling the story. The story world is thus located outside the time and place of the narrator. Imitation on the other hand is the mimetic enactment of the story. Here the story is told through dialogue, speech or bodily gestures from within the story world.¹¹

The term diegesis has reappeared a number of times in literary analysis. Gérard Genette re-appropriates the term making a distinction between the story told and the diegesis. The diegesis, he explains, rather than being the story itself, is the universe in which the story takes place.¹² Karol Berger addresses this discussion outside the context of literary analysis, considering the diegesis of visual and auditory events in paintings and music. Berger argues that we may connect the visual expression of paintings to

7 Nicholas Cook, "Methods for Analysing Recordings," in *The Cambridge Companion to Recorded Music*, ed. Nicholas Cook et al. (Cambridge: Cambridge University Press, 2009), 243.

8 Simon Trezise, "The Recorded Document: Interpretation and Discography," in *The Cambridge Companion to Recorded Music*, ed. Nicholas Cook et al. (Cambridge: Cambridge University Press, 2009), 207.

9 Gérard Genette, *Narrative Discourse Revisited* (Ithaca: Cornell University Press, (1983/1988).

10 Gorbman, *Unheard Melodies*.

11 Plato, *The Republic* (English translation by Paul Shorey) (London: William Heinemann Ltd., 380 BC/1930).

12 Gérard Genette, *Narrative Discourse Revisited*, 17.

the literary “voice.” Analogous to the way the “voice” communicates literary content, human figures in paintings communicate through visual bodily gestures presented to us in specific environmental settings.¹³ Similarly, in music, a vocalist may present a story world from within the perceived reality frame by, for example, singing about present feelings. He/she is then a character in the story world (mimesis). Or else the singer may present a story about another time and place. He/she is then the narrator of the world (diegesis). Also, as is often the case, the singer may switch between the different modes of presentation throughout the song.¹⁴ While there are obvious connections between literary narration and vocal utterance in music, there are less obvious, yet still significant, ways that a diegesis is established in the making of musical recordings.

Recorded tracks can be approached using the performance metaphor – as virtual performances,¹⁵ perceived performances¹⁶ or phonographic performances¹⁷ – and this line of thinking conditions the argument in this article as well as the aesthetic reflections presented in a vast number of books and articles dealing with analysis and production of music recordings. The performance metaphor is contingent on a causal listening mode. Here the listening experience is directed towards the behavior and gestures of performers associated with the production of the musical piece. These gestures belong to the sound-producing event as the way the performers express themselves. Simon Frith uses a visual analogy claiming that:

(...) to hear music is to see it performed, on stage, with all the trappings. I listen to records in the full knowledge that what I hear is something that never existed, that never could exist, as a ‘performance’, something happening in a single time and space; nevertheless, it is now happening, in a single time and space: it is thus a performance and I hear it as one.¹⁸

Although the presence of gestural elements in auditory experience is largely dependent on the particular genre (the performative aspect of the listening experience is definitely stronger in a Rolling Stones track than in one of Brian Eno’s ambient productions), the exploratory dimension of auditory perception¹⁹ is integral to most music listening. In the rock domain this link is particularly salient where the tracks often construct the idea of a distinct performance event, that is, the illusion that the music was performed simultaneously despite layers of non-simultaneous recording takes that have

13 Karol Berger, “Diegesis and Mimesis: The Poetic Modes and the Matter of Artistic Presentation,” *The Journal of Musicology* 12, 4 (1994): 407-433, 415.

14 For a thorough account of self-impersonation see Wendy Doniger, *The Woman Who Pretended to be Who She was: Myths of Self-imitation* (Oxford and New York: Oxford University Press 2005).

15 Allan F. Moore and Ruth Dockwray, “The Establishment of the Virtual Performance Space in Rock.” *twentieth-century music* 5, 2 (2008): 219-240.

16 William Moylan, *The Art of Recording: The Creative Resources of Music Production and Audio* (New York: Van Nostrand Reinhold, 1992).

17 Cook, *Methods for Analysing Recordings*.

18 Simon Frith, *Performing Rites: On the Value of Popular Music* (Oxford: Oxford University Press, 1996), 211.

19 Eric F. Clarke, *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning* (New York: Oxford University Press, 2005).

gone into its making. The discussions in this article is limited to recordings well-suited to the performance metaphor (i.e. recordings that encourage us to listen to it as a performance event framed by a particular time and place) in order to argue for some of the extended meanings that emerge from the staging of sound in the recording studio.

Taken as perceived performances, recordings are affective because they allow the listener to engage in an imagined relationship with the performer, by participating in a perceived performance event. Philip Auslander has called this affective state, the experience of *liveness*, by arguing that in spite of the fact that recordings represent a temporal gap between the performance and the reception, recordings form a technologically mediated co-presence of a performance.²⁰

One may argue that there is always a diegetic layering in record listening; this is, in fact, true. What we witness when we put on a record is the medium telling the story of a performance. This is no different from the relationship between the videotape and the film. The record, as well as the videotape, underlines in a very direct way the temporal and spatial gap between the perceived event and the moment and place of play back. It is through the recordings “act” of playing the event back to us that we come to witness the performed event, and it is here that we are presented with the ontological division between the “here and now” of the playback situation and the “then and there” of the perceived performance.

When we listen to music we have before us a phonogram – either in the form of a physical media (e.g., vinyl record or CD) or as a media-file (e.g., mp3 or wav). In a certain sense, phonograms are to record producers and sound engineers, what books are to writers. There is an author (or several authors) behind the specific configuration of sounds presented to us in recordings; this is analogous to the authors behind every word we read in a book. Despite these similarities, few people buying music would probably know the record producer involved in the recording; in contrast, most readers will know the author behind the book. Books, also, are usually listed by author’s name, while popular music phonograms are seldom, if ever, displayed by the name of the sound engineer or producer (unless the engineer or producer equals the performer). In this way, the musical performer certainly takes center stage in the history of musical recordings.

This fact, however, only tells us something about sound engineers’ hidden status in the history of record making, not, as a number of scholars have acknowledged, that sound engineers have actually played a profound role in the way music records have come to sound.²¹ While it may be contentious to entitle record producers authors (or co-authors) of the recording, it allows us to focus more on the way sounds are staged within the virtual space of the recording. Record producers mold sounds in and after the recording and thus create imaginary images of performances for the listener. In this way, recordings present a narrative emerging from the specific structure of sound events, analogous to the narrative present in books that emerges from the specific

20 Philip Auslander, *Liveness: Performance in a Mediatized Culture* (Oxon: Routledge, 1999).

21 see Mark Cunningham, *Good Vibrations: A History of Recording Production* (Chessington, Surrey: Castle Communications, 1996), and Albin Zak, *The Poetics of Rock: Cutting Tracks, Making Records* (Berkeley: University of California Press, 2001).

structure of words – presenting a specific angle the story the specific way the story is told. Books and phonograms are the final product presenting the story or perceived performance in its fixed form. In this way, one of the tasks of record producers is to present a specific story²² of a performance among the many that could be told.

From an audio-visual to an auditory diegesis

To outline this argument further I want to propose an analogy between filmic events and sound events in recorded music. The events we experience in films are that of characters acting out their lives. These events frame the film's fictional reality – its diegesis. The diegesis of the film is thus the presented filmic "reality;" that is, everything that logically belongs to the fictional world the film portrays, whether included (on-screen) or excluded (off-screen) from the picture frame.²³ On-screen sounds are, for instance, voices of characters on the screen or music coming from visualized playback systems. Off-screen sounds can be voices of characters that are not visualized, environmental sounds or music that belongs to the world of the characters, that is, music played back from a playback device outside the screen but in the film's fiction.

Extra-diegetic sounds emerge from outside the filmic reality; for instance, background music or other sound effects added to bring the viewer into a certain mood. The sources of non-diegetic sounds are not visualized, such as the voice over often heard in documentary films. Claudia Gorbman proposes that sounds may also operate at a meta-diegetic level, meaning that sounds belong within the character; as is the case, for example, with sound presented as belonging to the character's imagination.²⁴

Bordwell and Thompson present a more overall distinction between non-diegetic (non-story space) and diegetic (story space). These categories are further sub-classified in terms of the temporal relation between sound and image. Diegetic sounds may occur either simultaneously or non-simultaneously (earlier or later in the story) relative to the image, such as in the case of sound flashbacks or flash-forwards, while non-diegetic sounds may be marked as past (e.g., historical voice over present images), present (e.g., a narrator speaking in present tense) or later (e.g., a reminiscent narrator). Bordwell and Thompson state, however, that the temporal relationship of non-diegetic sound to the story is mostly irrelevant to the filmic experience.²⁵

In recorded music the perceived temporal relation between sound events has a more profound role for the classification of diegetic, meta-diegetic and extra-diegetic

22 Since this article deals with virtual performances as something constructed in the recording studio, I take story and narrative to be interchangeable terms. In this way, I see no difference between how the story is told (the narrative), and the perceived event (the story), both are products of narrating (see also Genette, 1983/1988). A study of recordings as documents of actual events should obviously distinguish between narrative and actual story.

23 Gorbman, *Unheard Melodies*; Christian Metz made a similar distinction between what he called 'ordinary sound' and 'sound-off', see Christian Metz, *Language and Cinema* (De Gruyter, 1974).

24 Gorbman, *Unheard Melodies*, 22-23.

25 David Bordwell and Kristin Thompson, *Film Art: An Introduction* (New York: The McGraw-Hill Companies), 1997.

events. Meta-diegetic sound events are indexes of events that took place before the diegetic events, while extra-diegetic sound events are indexes of events that take place after the diegesis. As already stated, contrary to film, diegetic layering in recorded music is not conventional practice. When the experience of extra-diegetic or meta-diegetic events occurs, however, they often emerge, as we shall see in the next section, as a consequence of both a spatial and temporal hierarchy between sound events.

The virtual space of recordings

Record listening is an instance of what Pierre Schaeffer called an acousmatic situation²⁶, that is, a situation in which we hear something without seeing its cause. According to Schaeffer this situation may potentially (or ideally) lead to a listening experience stripped of any relation to a sound-producing event (reduced listening).²⁷ Several other scholars, however, argue that auditory perception is fundamentally exploratory.²⁸ When listening to sounds we seek information regarding what is going on — what is that sound the sound of? What kind of event is taking place and what kind of source is involved? Michel Chion even claims that the acousmatic situation, rather than removing causal references, actually enhances the listener's interest in sound events.²⁹ In this way the acousmatic situation results in a new and, in many ways, enhanced aesthetic involvement compared to actual musical performances.

In audio-visual media the question of diegesis emerges in the relationship between sound and image. In recordings the question of diegesis emerges in the relationship between sound events and other sound events, more particularly in the relationship between sound events and the imaginary image of the performance. This perceived performance takes place in an imaginary performance setting that is most often conceptualized in terms of its impossible³⁰ or unnatural³¹ characteristics compared to actual acoustic spaces. This space is dimensional although often not with clearly defined acoustical properties.

In the most general sense we may think of space, whether actual or virtual, as either a prerequisite or a product of perception; that is to say that in every act of perceiving something, space is presupposed. Like the scene in a film, the auditory performance takes place somewhere. Sound sources are presented to the listener in certain spatial configurations that form a performance space. William Moylan has proposed the con-

26 Acousmatics is derived from the name Pythagoras gave to his uninitiated students (*akousmatikoi*), who received their lectures from behind a veil.

27 Pierre Schaeffer, "Acousmatics," In *Audio Culture: Readings in Modern Music* eds. C. Cox and D. Warner (New York: Continuum, 1966/2004).

28 See Clarke, *Ways of Listening*; Simon Emmerson, *Living Electronic Music* (Aldershot: Ashgate, 2007); Luke Windsor, "Through and Around the Acousmatic: The Interpretation of Electroacoustic Sounds," In *Music, Electronic Media and Culture* ed. Simon Emmerson (Aldershot and Burlington: Ashgate, 2000).

29 Michel Chion, *Audio-Vision*, 32.

30 Timothy Warner, *Pop Music: Technology and Creativity: Trevor Horn and the Digital Revolution* (Bodmin: Ashgate Publishing Limited, 2003).

31 Francis Rumsey, *Spatial Audio* (Oxford: Focal Press, 2001).

cept of the sound stage to describe the perceived two-dimensional stage on which sound sources are located. According to Moylan the sound stage is located in front of the listener at a certain distance.³² Sound sources often have different locations on the sound stage, and often they exist in different acoustical spaces; yet, listeners will unconsciously group them together to form a coherent performance.³³

It is tempting to define the sound stage as a diegetic frame in so far as the sound stage – the specific spatial formation formed in the act of listening – often frames the perceived performance. Being the focal space of the listening experience, the sound stage is, however, more like the visual screen in films. For instance, we may think of sound events that appear to arrive from outside the sound stage, yet still “participating” in the perceived performance. Such examples are clearly considered exceptions and often involve quite radical editing. We may imagine, for example, hearing a sound event with a perceived location in an adjacent room, performing with another performer located just in front of us. The argument here is that while everything located on the sound stage must belong to the diegesis of the perceived performance, everything that belongs to the performance need not be located on the sound stage.

My use of the word performance is understood as an event that takes place in a given time and place. The perceived performance may, as is often the case, include all sound events in the recording. However, in some cases sound events exist outside the temporal and spatial frame of the perceived performance. Clearly, there can be no sound events in recordings without a virtual space in which the listener can situate these events. We may, however, say that the recording holds a single diegesis, only in so far as all sound events in its virtual space are in fact perceived as being temporally related. If not, we may think of some sound events as belonging to an extra-diegetic or meta-diegetic level.

The phonographic diegesis 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. Recordings carry a virtual history as well as a virtual space. While the perceived performers may be heard as performing at the same time in the same place, they may also be heard as performing before and after each other. I am not referring to musical linear time, but to virtual historical moments where we perceive that the events took place. In this way, there is a very direct connection between virtual spatiality and virtual history, as the spatial separation is often realized in terms of the temporal separation between sound events. We may exemplify this with Queens of the Stone Age’s “You Think I Ain’t Worth a Dollar, But I Feel Like a Millionaire” mentioned in the introduction of this article. In this recording, the change in the perceived point of audio is not only related to spatiality (e.g. the performance leaps from one perceived performance space to another); the changing point of audition is also related to temporal change (e.g. the performance leaps from a certain point in the recordings virtual history – something pre-recorded played on the radio – to a later point in time – the moment of playback in the car setting).

32 See also Justin Christensen, *Virtual Distance and Soundstage, and their Impacts on Experienced Emotional Valence*. page 101-118 in this issue.

33 Moylan, *The Art of Recording*.

This leads me to my definition of diegesis in recorded music. The diegesis equals the perceived temporally connected performance. Sound events in the diegesis simply take place at the same time and the perceived performers can then logically respond to (hear) each other. In some cases, the diegetic bond is broken. This happens, for instance, when we experience that perceived performers respond to pre-recorded material (meta-diegetic sounds) or sound events that emerge after the perceived performance took place (extra-diegetic sounds).

Arguably, listeners tend to group the sound events into a single overall stream.³⁴ It simply makes sense to listeners to hear recordings as perceived performances, despite the spatial paradoxes often involved in doing so. Breaking up sound events into ontologically different layers is, therefore, something that demands radical recording techniques and sound editing, as I will exemplify further in the second part of this article.

Classification of sound events and the creation of meaning in recordings

As already proposed, certain ways of categorizing filmic sounds are applicable in describing certain kinds of phonographic sound events and to account for the creation of meaning in music production. In most recordings all sounds are grouped together on a single sound stage, forming a unified perceived performance. There are, however, several examples of tracks that do not present a single perceived performance space and thus challenge the idea of the sound stage. I propose four categories (Table 1) that describe the relationship of sound events to the perceived performance space.

Diegetic on-stage sound events:	Sound events in the perceived performance space.
Diegetic off-stage sound events:	Sound events absent from the perceived performance space, but temporally present.
Meta-diegetic sound events:	Internally re-produced non-diegetic sound events. These sound events takes place prior (in the virtual history) to the perceived performance, reproduced from a media within the sound stage.
Extra-diegetic sound events:	Externally produced non-diegetic sound events. These sound events takes place after (in virtual history) the perceived performance, causing the perceived performance to appear reproduced from within a new embracing ontological frame.

Table 1: Classification of phonographic sound events

Diegetic on-stage sound events are the most common category. Although musico-logical analysis may reveal that a track consists of many different acoustical environments, the listening experience most often involves hearing the perceived performance

34 Albert S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge and Massachusetts: MIT Press, 1990).

as spatio-temporally present in spite of the apparent acoustical paradox involved in doing so.

Diegetic off-stage sound events are defined as sound events that appear acoustically non-present in the primary acoustical environment. These sound events “participate” in the perceived performance events and are thus temporally co-present, however, the off-stage sound events are acoustically separated from the on-stage events. This acoustical separation can involve illusions of sound events located behind a physical obstacle, for instance, sound sources located in an adjacent room.

Meta-diegetic and extra-diegetic sound events are examples of a layering of perceived performances. In both cases one sound universe is presented through the other. Hence, two juxtaposed performances merged together in the track, for instance, as is the case in mash ups, are not understood here as an instance of diegetic layering. Diegetic layering, in my definition, emerges when listeners get the illusion of a hierarchical relation between two virtual spaces. Meta-diegetic sound events are sound events emerging from a media within the diegesis of the perceived performance. Analogous to paintings that portray a room in which another picture is present,³⁵ meta-diegetic sound events are illusions of pre-recorded sound events played back in the perceived performance situation. Logically performers (on the sound stage) can hear these sounds while performing. There is thus both a temporal and spatial division between diegetic sound events and meta-diegetic sound events.

Extra-diegetic sound events emerge outside and after the diegetic performance. Like meta-diegetic sound events, extra-diegetic sound events emerge as a media-within-media effect. However, in the first case, the meta-diegesis is in the embedded media, while in the second case the diegesis is in the embedded media.

Although the categories I propose are easily definable, they are not always straightforward to use in the description of recorded sound. Problems arise from the fact that the diegesis is not always clearly defined in listening, and from the fact that sound events do not necessarily have the same ontological status throughout a track. In this way, the ontological hierarchy of sound events may change from the first hearing of a track to the second, as the scene of the performance is formed in the listening situation.

Off-stage sound events

Filmic off-screen sounds are sounds with non-visualized sources that take part in the filmic events. In the case of the voice, we would define it as off-screen when we, the audience, cannot see the mouth of the filmic character speaking.³⁶ Phonographic off-stage sound events are defined in terms of the spatial (acoustical) separation between sound events. In films there need not be any acoustical division between on-screen and off-screen sounds. In fact off-screen sounds may often appear acoustically

35 Berger, “Diegesis and Mimesis,” 421-422.

36 Michel Chion and Claudio Gorbman, *The Voice in Cinema* (New York: Columbia University Press, 1999).

closer to us than on-screen sounds. Phonographic off-stage sounds on the other hand emerge when we experience that performers are located in different environments. As a result they cannot “see” each other. In this way, phonographic off stage sounds are defined from the point of view of the performers. What makes off-stage sound events comparable to off-screen sounds is that off-stage sound events mimic sound sources that would not be visible if we could see the performance. Thus, in order to account for the link between off-stage sound events and filmic off-screen sounds, we have to imagine what we would be able to see in an actual performance.

A notable example of off-stage sound events occurs in the tune “I Won’t Hurt You” from the album *Smiling and Waving*³⁷ by the Norwegian singer Anja Garbarek. In this track, there is a significant acoustical separation between the lead vocal in front and a guitar and percussion sound. Although the percussion sound appears too mechanical to be actually performed, the environmental co-presence with the humanly performed guitar assists in animating the percussion sound.³⁸ The off-stage effect occurs primarily as a result of low-pass filtering around 400Hz (with a fairly steep slope) on guitar and percussion. In this way, the listener may experience some concrete material separating percussion and guitar from the sound stage. Accordingly, we get the impression that we would be able to see the vocalist, who is located just in front of us, while the sounds heard emanating from behind the barrier are invisible. For this reason we may think of the vocal as on-stage and guitar and percussion as off-stage. Following this line of thinking the guitar sound and the percussion sound appear inaccessible to both the listener and the on-stage vocalist and therefore signal a spatial rupture in the horizontal dimension. The vocalist appears lonely on the sound stage, yet she participates in a shared performance with off-stage sound events. This experience emerges as a result of the combination of a temporal co-presence and an acoustical separation. The example, then, illustrates the importance of distinguishing between a sound stage (a circumscribed acoustic space) and a perceived performance universe (the totality of synchronous sound events).

Contrary to filmic off-screen sounds that may change to on-screen sounds if the source becomes visible,³⁹ phonographic off-stage sounds are not loaded with such expectations. We simply do not expect phonographic characters to move. Neither do we expect a change of listening perspective.

The experience of off-stage sound events does not necessarily occur in all tracks where a closer analysis may reveal acoustically detached sound events. Most of the time, listeners group sound events into a single perceived performance in spite of the inherent spatial paradoxes that this grouping implies. Off-stage sounds often emerge

37 Anja Garbarek, “I Won’t Hurt You,” *Smiling and Waving* (Virgin, 2001).

38 We may also notice the semiotic connotations of the percussion sound, relating it to the sound of the performer’s heartbeat. The sound may, then, be heard as something internal to the performer – something the listener hears only through the vocalist hearing, feeling or imagining it. Since the guitar sound is located in the same acoustical environment as the percussion track, we may also locate the guitar sound in the mental interior of the performer. In this sense, we may think of the percussion and guitar sounds as meta-diegetic – on-stage and internal – instead of off-stage.

39 Chion, *The Voice in the Cinema*.

as a result of heavy processing, for instance applying a high cut filter at 250 Hz or other spatial effects that clearly creates the illusion of two separate environments. Whether this illusion in fact occurs, obviously depends on many aspects related to both production and reception. While the visual frame in film clearly separates inside and outside, such borders do not (or only vaguely) exist in recorded music. For this reason, we must be content with blurred lines between phonographic off-stage and on-stage sound events.

Meta-diegetic sound events

While it may prove difficult to argue for off-stage sound events in recordings, one can find many clearer examples of meta-diegetic sound events. Pink Floyd's "Wish You Were Here" from the album *Wish You Were Here*⁴⁰ from 1975, for instance, has a very salient diegetic layering of sounds.

The track opens with the sound of a radio being tuned. As the tuning settles on a station, a guitar sound is heard, apparently emerging from the radio. At this point, a diegetic layering is already implied between the radio playback and the performance within the radio. We do not, however, hear any acoustical signs of a virtual playback environment in which the radio, logically, must be situated. The sound from the radio simply does not activate an acoustic space and we do not hear any signs of human presence "outside" the radio. Thus, the radio in-itself remains an abstract non-spatial object. This experience is particularly strong when played back through headphones, where the radio can only be heard through the right headphone.

This experience changes quite radically when we hear coughing and breathing noise in the center position at 0'26. The coughs and breath noises do not arrive from the radio, rather, from the room in which the radio is playing. In this way, we come to experience a perceived radio listener and, accordingly, the spatial and temporal separation between the radio listener and the radio performance. The diegetic layering of sound events is, thus, fulfilled as we hear both the radio performance and the listener simultaneously. As a guitar starts playing at 0'42, we get the illusion that we are hearing someone playing along to the radio track. As a result, the playback environment is transformed into a combined playback and performance environment. We hear the radio performance in the knowledge that the guitar performer hears it too. The fact that the diegetic frame is consolidated gradually on top of the meta-diegetic level seems to intensify the presence of the diegetic frame. This effect emerges as a result of the shift in auditory focus. Listeners are likely to naturalize the meta-diegetic level in the first bars of the radio-guitar – attend to it as if it was diegetic – only to be surprised by the presence of the guitar when the diegetic level is set in motion.

At 1'34 the vocal enters. From this point in the song, the radio performance gradually becomes less noticeably, and at the 1'36 it disappears completely, without any clear demarcation of its disappearance. As the drums enter, followed by bass and

40 Pink Floyd, "Wish You Were Here," *Wish You Were Here* (EMI, 1975).

piano, the performance environment is transformed (expanded) again into a stage. In this way, the intro presents a gradual shift from a performance in an “everyday setting” to a performance “on stage.”

The intro of “Wish You Were Here” is an example of a media-within-media effect, or to expand on William Moylan’s terminology,⁴¹ a sound stage in sound stage experience (Figure 1). The diegetic guitar performer situated in Sound Stage A can hear the meta-diegetic guitar performance in Sound Stage B. On the opposite, the performer on Sound Stage B cannot hear the performer on Sound Stage A. In other words, it is the diegetic performance that makes it possible for us to hear the meta-diegetic performance, and not the other way around.

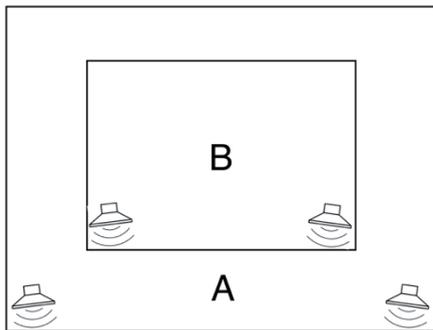


Figure 1: media-within-media

While it makes sense to talk about an ontological rupture between sound stage “A” and “B,” the addition of sound stage “A” have implications for the way listeners attend to sound stage “B.” The primary perceptual effect pertains to a further mediatization of sound stage “B.” The effect of another layer of virtual mediation is a weakening of the indexical relation to a pre-medial world. It is not that the performance on sound stage “B” necessarily ceases to be heard as one. It is rather that the perceived performance recedes into the background, as a mediatized phenomenon in sound stage “A,” something that is available to us only through it being present to someone else.

The media-within-media effect is in fact a common type of media based staging⁴² heard in many different forms in newer pop and rock albums. “Wish You Were Here,” however, is a particularly clear case as the performance space is gradually established. From the outset we get to hear a performer on the radio, what we may call a meta-diegetic performer (analogous to Genette’s meta-diegetic characters⁴³). While meta-diegetic sound events are still indexical of meta-diegetic performers, the index is generally weaker compared to diegetic sound events. The meta-diegetic performer appears to us, only as the last part of the indexical chain, embedded in another index and on the wedge of slipping from the listening experience.

41 See Moylan’s definition of the sound stage in *The Art of Recording*.

42 Simon Zagorski-Thomas, *PhD in Music Composition*. PhD Thesis (University of London, 2008).

43 Genette, *Narrative Discourse Revisited*.

Another interesting case of diegetic layering is found in Björk's "There's More To Life Than This" from her album *Debut*.⁴⁴ The track opens with the sound of a lot of people chatting in a large room. A dance beat emerges in the background and we realize that the setting is some kind of dance club. At 0'35 Björk's vocal is heard close-up. Throughout the track the listener is locked to this position just in front of Björk as she walks around in the dance club. Everything the listener hears is from Björk's point of audition or more specifically from the position just in front of Björk's mouth. In this way, the track has a dynamic sound stage analogous to a moving camera that gradually changes what is considered "on-screen." Like point of view screen shots in films,⁴⁵ the point of audition⁴⁶ perspective in "There's More To Life Than This" generally enhances our identification with the vocalist.⁴⁷

The sound of the background music seems to arrive from a set of loudspeakers in the room in which Björk is positioned. Accordingly, we may think of the background musicians as located on a meta-diegetic level. The music in-itself is of course diegetic in the sense that Björk can hear the music in the room she is in. However, this is not the case the other way around. The musicians performing the background music do not share reality with Björk. Thus, the background music stands in a hierarchical relation to Björk's performance in that the ontological level of the former mediates through the ontological level of the latter.

At 1'37 we hear Björk open a door to enter an adjacent room, and as she closes the door behind her, background sounds are cut off. In this way, we get the experience that Björk is now alone in another room and therefore not visibly connected to the rest of the sound sources (the sound system and the chatting people) in the recording. There is, thus, a clear division between on-stage sounds (Björk's voice) and off-stage sounds in this scene. Björk's intimate vocal address is further intensified⁴⁸ due to the one-to-one illusion provided by the virtual environment.⁴⁹

We may think of the back track as doubly displaced. Since it emerges through a medium located off-stage, we may conceive of it as an off-stage meta-diegetic performance. As a result of the ontological division Björk takes on the role as the narrator addressing the background music. She is, so to speak, telling the story about the other

44 Björk, "There's More to Life than This," *Debut* (Elektra/Asylum, 1993).

45 James D. Buhler, David Neumeyer and Rob Deemer. *Hearing the Movies: Music and Sound in Film History* (New York and Oxford: Oxford University Press, 2010).

46 Chion, *Audio-Vision*, 89-92.

47 Philip Auslander – referring to Mary Hunter, "The Stage Director in Cinema", In *Theatre Arts* (May, 1949) – makes a similar comparison between a stage director's "manipulation of audience attention" and the moving camera lens in film that move from one point of interest to the next – a "psychological camera eye". Philip Auslander, *Liveness: Performance in a Mediatized Culture* (Routledge 1999).

48 See Nicola Dibben, "Vocal Performance and the Projection of Emotional Authenticity" In *The Ashgate Research Companion to Popular Musicology* edited by Derek. B. Scott (Farnham and Burlington: Ashgate, 2009).

49 The track was allegedly recorded live at the (now closed) Milk Bar nightclub in London. Using a handheld microphone Björk walks through the club to the toilet as the backing track is played back at the club's sound system, see Ian Gittins, *Björk: Human Behaviour: The Stories Behind Every Song* (London: Carlton Books, 2002).

ontological layer, and responds to its rhythmic and tonal structure. What is interesting is the way the causal relation between the background music and the voice seem to change direction through the track. Björk is clearly responding to the background music, as she sings along to it, and her performance is, thus, logically located after the events heard in the background recording. The choir voices heard dubbing her voice at 0'47, however, announce a sense of simultaneity, and the causal chain even reverses at 1'04 where the choir seems to respond to Björk's voice. Taking these events literally we are presented with a spatial and temporal paradox, arising from the experience of perceived performers responding to events that takes place after the moment of responds.

Extra-diegetic sound events

Extra-diegetic sound events refer to anything "outside" the performance universe, that is, anything the performers on-stage cannot hear. We find an example of this in Air's "Radio #1" from their album *10.000 Hz Legend*.⁵⁰ At 3'14 we suddenly hear a male voice humming along to the chorus. The male voice sings along to the track and, thus, takes part in the performance. He appears, however, temporally separated from it. When the voice enters, the listener suddenly hears the initial performance through another diegetic layer. In this way the listening perspective changes. This is achieved by ducking the diegetic sound events with side chain compression when the male voice enters. This effect is common in many documentary films, television commercials, and in radio broadcasting when the speaker presents a track, while we hear the music in the background. In television these voices are called voice-overs. In the purely auditory domain they do not have a name, but as they are clearly comparable to the voice-overs we encounter in audio-visual media, phonographic voice-overs seems an appropriate term here.

Considering the title and the lyrics of the song, the impression of a radio speaker humming along to the track immediately comes to mind. There are, thus, two temporally separated perceived performances: (1) The radio speaker's performance that enters at 3'14; and (2) the performance formed by the rest of the sound sources. The radio speaker is addressing the diegetic performance in a very direct sense by singing along to it. He is, in one sense, pointing us in the direction of the song; in another sense it can be perceived as an intrusion, interrupting the relationship between the performance and the listener. The listener must zoom out when the radio speaker enters to embrace both performances. In this way, the diegetic performance seems more distant in time and space when the extra-diegetic voice is forced upon the listener.

Where Air's "Radio #1" plays with an extra-diegetic voice, the Japanese noise rock band Boredoms plays with extra-diegetic medium sounds in the track "Super You" from the album *Super AE*.⁵¹ The track has no clear melodic or rhythmic structure, but consists of layers of distorted guitars floating in an out of each other. This noise atmosphere is disrupted several times during the track by the sound of a tape reel being

50 Air, "Radio #1," *10.000 Hz Legend* (Virgin, 2001).

51 Boredoms, "Super You," *Super æ* (Warner Music, 1998).

fast-forwarded several times – as if someone else were winding the entire track for us. The effect is most profound in the following intervals 0'30-0'32, 0'55-1'00, 3'31-3'51 and 5'16-6'30. Several clearly audible tape splices further enhance the impression of a tape recording. The tape-reel effect may be seen as an instance of diegetic layering. The tape-machine is not an effect in the performance, but a medium that belongs to another ontological world than the perceived performance. Contrary to the extra-diegetic layer in "Radio #1," however, the extra-diegetic layer in "Super You" lacks human presence. Instead the tape machine may be seen as a substitute for the human narrator. It is essentially the tape-machine that makes it possible for us to hear the performance – it tells the story of another time and place. The medium-within-medium effect found in "Super You" is interesting precisely because it changes the listener's otherwise direct relation to the performance through the recording, by adding an extra ontological layer between the listener and the performance.

Discussion

We are now in a position to return to the Queens of the Stone Age example mentioned in the beginning of the article. In "You Think I Ain't Worth a Dollar, But I Feel Like a Millionaire" we are presented with two listening perspectives: (1) through the position of the car driver hearing music through the car radio and (2) as a "directly" mediated performance presented to us through our own loudspeakers. Like this, we hear movement from a meta-diegetic to a diegetic level. The meta-diegetic level "eats up" the diegetic level, enhancing the experience of loudness and intensity. While we may say that the song is presented to us at a meta-diegetic level – as a consequence of its location in the car radio – the meta-diegetic level quickly obliterates as the quality of the sound changes, removing the car setting from experience.

While the physical division between sound and its natural source – the possibility to move sound to another host – became reality with the invention of the phonograph in 1876, the possibility to layer sounds, and potentially create diegetic layering of perceived sound events, emerged with sound-on-sound recording several decades later. After some of Les Paul's first experiments with sound-on-sound recording, Patti Page working with Mitch Miller on the Mercury Record Label used this idea to sing along to herself on her 1947 recording of "Confess."⁵² This recording is the first known example of sound-on-sound recording using tape, and was marketed as "Patti Page and Patti Page," suggesting how this effect was seen, at least partly, as an act of vocal gimmickry.

Of course, the emergence of multitrack technology in the 1950's quickly initiated many examples of singers dubbing themselves, and today the practice is so conventional that most listeners will not even notice the overdubs that have gone into making the many layers of voices they are presented with on different recordings. Listening (back) to Patti Page's recording with today's ears, there is, however, something dis-

52 Patti Page, "Confess" (Mercury, 1947).

jointed about the perceived performance caused by the sonic quality of the dubbed backing vocal. The backing vocal appears temporally and spatially displaced. The effect is unintentionally created and results from the audible technological artifacts – such as tape hiss – arising when recording from one tape to another. Yet, listeners may experience a media-within-effect placing the backing track on a meta-diegetic level.

In some instances, knowledge about the original event behind the perceived performance may have an effect on the listening experience. Such is the case with Natalie Cole singing a duet with her late father Nat King Cole on “Unforgettable.”⁵³ Although there is no clear indication in the actual sound that Natalie Cole’s voice was recorded 30 years after Nat King Cole’s voice, our knowledge about the making of the song may intrude on the listening experience. Natalie Cole is mediating the past event being both listener of her father’s recording, while performing to it and, in effect, creating a new one. In so far as we may mark Nat King Cole’s voice as a past event, his voice is, in a very peculiar way, both diegetic and non-simultaneous. Although the contextual story accompanying the virtual duet is noteworthy, the temporal separation between recorded takes is in no way uncommon. We may, for instance, think of rap vocals heard performing to sampled phrases with a clear origin in past recordings. As in “Unforgettable,” the temporal separation in rap music is mostly not an audible feature of the sonic material, but rather contextual information that impacts on the listening experience.

Conclusion

In this article I have suggested how the concept of diegesis may work to classify sound events in recorded music and how the concept is useful to account for the production of meaning in music production. With reference to Genette’s classification of narrative levels,⁵⁴ I have argued that diegetic layering of sounds within the virtual space of music recordings can be classified as meta-diegetic, diegetic or extra-diegetic. Although the discussed tracks shows examples of unusual spatial distributions of sound events, they support the argument that it makes sense to think of a phonographic diegesis that can be exploited creatively in music production to create different forms of narrative effects.

The idea builds on the performance metaphor and the belief that listeners hear recordings as documents of a fictional past event. I have argued that diegetic layers in recordings should be defined primarily according to the perceived temporal relationship between sound events. Meta-diegetic sound events are either performer-internal-sounds or pre-recorded performances played back through a media in the diegesis. Whereas film sound theorists often pay most attention to the former, I redefine meta-diegesis to account for the experience of a media-within-media effect. The notion of meta-diegesis then applies when we experience that diegetic performers (pres-

53 Natalie Cole, “Unforgettable,” *Unforgettable, With Love* (Elektra, 1991).

54 Gérard Genette, *Narrative Discourse: An Essay in Method* (Ithaca and New York: Cornell University Press, 1972/1980).

ent events) can “hear” meta-diegetic performers (past events), but not the other way around. Likewise extra-diegetic sounds emerge as a media-within-media effect. Extra-diegetic sound events, however, appear after the diegesis and thus logically become the new present as the diegesis is relocated to the past.

As I demonstrated through the examples the, diegetic layering of phonographic sound events may take various forms, and often the listener is presented with ambiguous or even conflicting acoustical cues. In addition, I suggested that previous knowledge about the production of the track have an effect on the nature of the perceived performance. In this way, the notion of the phonographic diegesis may open up for further reflection on the experiential aspects of virtual spatiality in recorded music.

Abstracts

The use of studio-based effects and recording techniques to shape the sound quality and aesthetics of recordings are addressed in a number of studies. This article explores how such practices have shaped the sonic narrative of recordings and listeners’ point of audition. Building on research into literature and film sound, and by considering the spatiality of a number of popular music productions, I outline the concept of diegesis in the context of recorded music. I conclude by proposing a way to categorize sound events relative to the perceived performance space of recordings.

En lang række videnskabelige publikationer har undersøgt den måde hvorpå optageteknikker og efterbehandling af musik i lydstudiet har formet optagelsens æstetik. Denne artikel har fokus på hvordan det auditive narrativ samt lytterperspektivet er designet i lydstudiet. Med afsæt i litteraturstudier og filmlydstudier, og igennem korte analyser af en række populærmusikproduktioner, argumenteres der for en redefinerings af diegese-begrebet og en kategorisering af lyd i optagelsens virtuelle performancerum.