The Wasteland Violin
Sonic Modality and Affective Immersion in
*Fallout 3: New Vegas*

Lauri Sallamo
Pro Gradu Thesis
English Philology
University of Oulu
Spring 2013
# Table of Contents

1. Introduction......................................................................................................3

2. Theoretical background and definitions............................................................5
   2.1. Immersion..................................................................................................5
   2.2. The soundscape..........................................................................................6
   2.3. Modality and coding orientations.................................................................9
   2.4. A note on types of modality........................................................................13
   2.5. Affect and expectation................................................................................14
   2.6. Distance and perspective............................................................................17

3. Corpus and analysis........................................................................................22
   3.1. Soundscape of the wasteland......................................................................23
      3.1.1. Sound effects....................................................................................24
      3.1.2. Music................................................................................................28
      3.1.3. Blend of sound effects and music.......................................................29
   3.2. Modality and affect in F3:NV.....................................................................30
      3.2.1. Diegetic sound effects.........................................................................31
      3.2.2. Non-diegetic sound effects..................................................................34
      3.2.3. Borderline diegetic sound effects.......................................................34
      3.2.4. Diegetic music....................................................................................36
      3.2.5. Non-diegetic and borderline music....................................................36
   3.3. Disconnect as agent of immersion...............................................................39

4. Discussion and conclusion..............................................................................43

Bibliography........................................................................................................45
1. Introduction

Games are in large part popular because of their ability to stimulate thinking, to pass time and to entertain. The most successful games are typically those that keep their players coming back; those that can consistently provide engagement, a string of positive stimuli and an internally logical alternate environment for players to identify with. In the context of video games, the state of being entertained and intellectually engaged is commonly referred to as immersion. The deeper the experience of immersion, the less likely players are to view the game world simply as a construct and more as a quasi-real space with actual real-life relevance. Different types of video games favour different types of sound design, but in my experience sound design has a surprisingly pivotal role in producing the experience of immersion regardless of game genre. Often the best soundscape is not the one a player is always conscious of but the one that blends in and fits the game and its aesthetic on a more subconscious level.

As mentioned above, sound events have different roles in different soundscapes, and for a positive emotive response to occur and for that response to strengthen the experience of immersion, they need to be appropriate for the contexts they are likely to be heard in. Each sound event needs a suitable context and an associated meaning. In the case of a video game, the context is always in part visual, but also temporal like in any artificial soundscape, taking into account what came immediately before the event and what is likely to come after. This study is concerned with the modality or 'truth value' of sound events, but high modality is not necessarily synonymous with realistic sound design. Rather, different contexts have different modal requirements for sound events for them to produce a desired effect in the player.

This study attempts to find new tools for analyzing video game soundscapes on a general, broader level, examining the mechanics underlying the onset of emotive response and consequently immersion. As the game under scrutiny I have selected *Fallout 3: New Vegas*, a post-apocalyptic first-person role-playing game published for the PC in 2010. While making the effort to assess the sonic modality and
affective mechanisms of various different sound events found in the soundscape crafted for Fallout 3: New Vegas, this study makes the presumption that higher sonic modality has a correlation with a deeper experience of immersion in the video game while acknowledging the likelihood that there are also exceptions to this rule. I will study the typical roles different types of sound events tend to fill in the game and the ways in which they relate to the idea of distance and perspective in its soundscape. Finally, I will discuss how sound events may combine, intentionally or otherwise, to produce either affective response and immersion or an adverse reaction in the player.
2. Theoretical background and definitions

I have chosen three works by as many authors as the theoretical foundation of this study. The three authors are music educator and environmentalist R. Murray Schafer, linguist and semiotician Theo van Leeuwen, and composer and philosopher Leonard B. Meyer. All three have distinguished themselves as central actors in their respective fields. In addition to outlining some of their theories that are salient for this study, I will discuss the concepts of immersion, modality and perspective on a more general level.

2.1. Immersion

Since immersion (and the subjective experience thereof) is both a desired result of audiovisual entertainment and the only evaluative gauge used in this study, it is essential to clarify what is understood by it in this context. While innately being a subjective experience and therefore a somewhat elusive concept, immersion could perhaps best be described as the level of attention and emotional engagement a video game is able to induce in its player. I would venture as far as to state that an experience of immersion propitious not only for a continued but a gradually deepening engagement in the game can be regarded as a universally desired goal in game design. Games of different genres go about achieving this goal in slightly different ways, but every video game is concerned with creating and maintaining a consistent, internally logical world for players to identify with, which in turn requires a soundscape that serves the purposes and needs of that particular game.

Grau (2003, quoted in Collins 2008: 133) defines immersion as being “characterized by diminishing critical distance to what is shown and increasing emotional involvement in what is happening”. Collins acknowledges that there are many possible and debatable subdivisions of immersion, for instance sensory immersion (forceful visual and aural input draws player's attention), challenge-based immersion (game requires skill and offers satisfying challenges) and imaginative immersion. She focusses on imaginative immersion – the type most commonly understood by
the term – in her book, as will I in the present study. Imaginative immersion stimulates creative thought processes and enables the player to empathize with game characters and enjoy the fantasy of the game. As the player becomes more emotionally invested in the story of the game, the more real the environment the story is set in becomes. Although visual and aural fidelity are important factors of immersion (their significance largely depending on game genre), Collins points out that the realism aspired to in games is in many ways not a naturalistic realism in the sense of being a simulation of reality, but a cinematic realism that relies on established motion-picture convention (2008: 134). Sounds can be metaphoric and meaningful on levels other than just as an attempt to approximate reality (2008: 135).

The level of immersion achieved during a particular gaming session is difficult to objectively measure, however, as immersion is essentially an individual experience subject to a plethora of personal preferences and dislikes. Gathering data about the experience of immersion, then, is dependent on the player's ability and willingness to verbalize said experience. Using myself as a test subject, I will utilize the method of self-reporting in this study and do my best to verbalize any experience of affective immersion that might take place during gameplay. Another way to gather data could be to monitor the player's heart rate, respiration and other physical indicators during a gaming session. However, aside from being somewhat less than a straightforward matter to implement effectively with my resources, it is likely that such analysis would not yield enough unambiguous data either, as similar affective experiences by different people may result in clearly differentiated physiological responses and vice versa.

2.2. The soundscape

The soundscape is a central term in this study, and I think it is appropriate to begin by looking into one of the foremost works of the person who coined it: *The Soundscape: Our Sonic Environment and the Tuning of the World* by R. Murray Schafer (1977). Schafer starts from the premise that the soundscape can realistically be any acoustic field of study – for instance a musical composition, a radio program
or an acoustic environment – consisting of “events heard not objects seen” (1977: 7-8). Despite its aural nature, “soundscape” fittingly alludes to “landscape”. Both contain the idea of relative and variable distances; of space and of perspective within that space.

Discussing perspective, Schafer separates sound events roughly into three categories, using terminology from the study of visual perception: figure, ground and field. Figure, also occasionally referred to as signal in the book, is a sound event requiring immediate attention and is typically consciously listened to. Examples of figure/signal could be an alarm clock or a dog barking, always depending on the context. Context is very significant because figure in one soundscape might easily be a keynote sound in another – i.e. a sound commonly found in a given environment, where its absence would be more likely to be noticed than its presence. There is a considerable difference, for instance, between a dog barking at home and a dog barking inside a kennel with a chorus of other canines. Ground is background ambient which is typically passively heard instead of actively listened to, while field is the area inside which figure, ground and the observer's relationship to them are defined. Field thus refers to the place where all the sounds occur, i.e. the soundscape (1977: 152).

Much like the perspective of a painting places the viewer in a desired position, so do aural dynamics situate the listener of any sound object – as occurs for instance with radio plays. Citing a three-stage plan of a radio technician, Schafer draws tentative parallels with the terminology presented above (1977: 157). A radio play sound scene consists of three main levels: immediate, support and background. The immediate effect is meant to be actively listened to, which makes it similar in function to a signal or foreground. Support effects occur in the immediate vicinity of the foreground, as do keynote sounds and ground, while background effects set the scene much like field does.

Schafer points out that aural experiences are also by definition tactile (1977: 11). This is true especially in low (bass and sub-bass) frequencies, where, passing below
20 hertz, audible sound can no longer be sensed by humans as other than vibration. He states that “touch is the most personal of senses” and that “hearing is a way of touching at a distance” (1977: 11), but goes on to emphasize that hearing is also a sense that cannot be turned off. Therefore, to protect our sanity and to help us enforce our bodily limits, we have adapted and developed defensive mechanisms in order to filter out unwanted noise in our environment. Sometimes actively concentrating on a task at hand or simply passively ignoring the offending sound can provide a sufficient sonic shield, but especially in noisy urban environments the offending sounds are often substituted with more pleasant aural stimuli. Schafer calls this phenomenon “audioanalgesia” (1977: 96), referring to the use of sound as painkiller. Many people find it easier to study or work while listening to music of their choice, masking potential interruptive signals in their immediate surroundings. Steady low-key ambient sounds like unintrusive background music in public spaces and even devices like air-conditioners can be regarded as instruments of audioanalgesia.

Schafer makes a distinction between hi-fi and lo-fi soundscapes, using as his primary example the transition from rural to urban environments. A hi-fi system, Schafer elaborates, possesses a favourable signal-to-noise ratio (1977: 43). This means that even quiet signals can be heard distinctly over long distances due to a low ambient noise level. A lo-fi environment, on the other hand, is very much the opposite: a signal often has to be amplified in order to be made out even across short distances. An example of a hi-fi soundscape could be the nocturnal hoot of an owl in a pastoral setting, while the din of a busy metropolitan street would certainly qualify as a lo-fi soundscape.

Schafer also makes an interesting connection between sound and power (1977: 51, 115). The church bell long dominated the soundscape of the countryside, its powerful tolling signalling a connection with the divine, but the Industrial Revolution and the continuing secularization of the western society has since shifted the power of “sacred sound” to other hands. Schafer expresses the idea that attending a rock concert may sometimes have surprisingly much in common with worshipping a deity.
and also points out that electric amplification ultimately negated the need for growing orchestra size in order to produce music of more impressive volume and dynamics. This also had interesting implications for the perceived intimacy of sound events, an idea I will return to below while outlining some of Theo van Leeuwen’s theories.

2.3. Modality and coding orientations

Theo van Leeuwen advocates treating speech, music and all other sound according to the same principles in his semiotics book, *Speech, Music, Sound* (1999). In his view, all sound events can be analyzed as agents for the communication of semiosis (sign processes), and the meaning potential of any sound event is dependent on the context it appears in. As he puts it, “the same sound can be used to mean one thing in one context and another in another context” (1999: 10). Or, conversely, the same meaning potential could be communicated using two completely different sounds in different contexts.

Central in Van Leeuwen’s treatise is the idea of perspective and social distance. He adopts Schafer’s terminology when discussing the relative levels of distance or prominence of sound events within a sound scene – i.e. figure, ground and field – with the proviso that they may also be symbolic places or positions and that there may sometimes be only two layers present, the foreground and the background (1999: 16). Van Leeuwen points out that these layers and their expected effects can easily be toyed with with the help of modern recording technology, and a shift in perspective can be used to force the listener to make new connections with a sound event that has previously been associated with another context. This applies, for instance, to perceptions of intimacy. A whisper, typically associated with close proximity to another human being, can be mixed to cut cleanly through a wall of sound, which under normal circumstances could not occur in a naturalistic soundscape. On the other hand, background ambient can be elevated in relation to signals to the effect that the figure begins to appear less relevant and/or more distant than the field it is supposed to be supported by. Van Leeuwen's example of this
phenomenon will be provided at the end of the analysis of my corpus.

Van Leeuwen discusses 'immersion' as the opposite of perspective. Although I will be using the term in a quite different sense in this study, it is nonetheless a fitting noun for the effect low-frequency diffusion has on a listener. While higher-frequency sounds (animal calls, for example) are easier to pinpoint and locate in space, lower frequencies spread more in every direction due to their broader wave length and are therefore more difficult to locate. When close to a powerful low-frequency sound source, the listener will in most cases experience being enveloped by the sound or being immersed in it. Incidentally, this physical sense of immersion in sound is quite often also important for the player's experience of affective immersion in video game soundscapes for reasons that will be discussed later in this study.

Van Leeuwen (1999: 180) defines modality in the context of sound as follows:

The term ‘modality’ refers to the degree of truth assigned to a given sound event. The term ‘truth’ takes on a somewhat different meaning depending on whether it pertains to presentation or representation. In the case of representation, 'truth' means 'a true representation of the people, places and/or things represented', in the case of presentation it means 'true to the spirit of the genre, and the values which underpin it in its context'.

As he points out, truth is not a static characteristic; rather, it is assigned to an object or event, and the modality of said object or event is then determined by how real or true it is (re)presented as. Providing an example, he states that the truth of an emotive statement lies not in whether the emotion is “really felt that way”, but in “as how emotional” it is expressed. Patrik N. Juslin (2009: 131) also makes a distinction between perception and induction of emotions, suggesting that it is possible simply to perceive an emotion in the music, or actually feel an emotion in response to the music. The modality value of a sound event, then, may be regarded as high even as the experience of an affective response may or may not be present. The onset of immersion requires an affective experience grounded in high modality, but high modality does not necessarily imply resulting immersion.
Acknowledging the different potential logics within various aural environments, Van Leeuwen discusses several kinds of modalities in terms of their coding orientations. He bases his discussion on previous work done on visual communication by himself and Gunther Kress (Kress and Van Leeuwen 1996) and goes on to apply its basic principles to the context of music and sound in general. Van Leeuwen states that “the modality value of a given modality configuration depends on the kind of modality which is preferred in the given context, and the choice of a particular kind of modality, in turn, derives from the coding orientation of that context, the values which underlie semiotic choice in that context” (1999: 160).

Van Leeuwen names articulatory parameters by which the modality of an image can be assessed: articulation of detail, articulation of background, colour saturation, colour modulation, colour differentiation, depth articulation, articulation of light and shade, and articulation of tone. He describes how amplified and reduced parameters can result in different modality configurations, which in turn will cue the observer's modality judgements (1999: 159). He goes on to outline four different coding orientations for visual contexts: naturalistic coding orientation, technological coding orientation, abstract coding orientation and sensory coding orientation.

Naturalistic coding orientation defines visual truth in terms of how much an object that is being observed resembles its real-world counterpart. If its resemblance is high, so is its modality and vice versa. Naturalistic coding orientation will usually prefer high articulation in detail and background and moderate articulation in most its colour parameters. (1999: 160)

Technological coding orientation, in turn, bases an image's truth value on its practical usefulness. Good examples are maps and blueprints, and their modality is high if they manage to fulfil their purpose in successfully guiding their user's movement and actions. Images of this type typically have reduced articulation for maximal effectiveness. (1999: 161)

Abstract coding orientation can be found, for instance, in modern art and scientific
visuals. Its basic premise is that “the more an image represents the deeper 'essence' of what it depicts, the more it represents the general pattern underlying many superficially different specific instances, the higher its (abstract) modality” (1999: 161). Reduced articulation is hence typical for abstract truth.

Sensory coding orientation, finally, is ultimately concerned with affect; the emotive effect an image can have on its viewer. Images with high modality of this type are characterized by more-than-real articulation of parameters, emphasizing colours, sharpness and depth, for instance. Instead of being informative or natural, first and foremost, images of this kind usually try to coax a reaction appropriate for the context they appear in. For example, food advertisements tend to have high detail and intense colour in order to whet the viewer's appetite, while perfume ads use imagery that evoke the sense of smell (1999: 162). As Van Leeuwen points out, synaesthesia is often a part of sensory modality.

Applying these theoretical principles to the modality of sound, Van Leeuwen provides us with three different coding orientations – abstract-sensory, naturalistic and sensory – and similarly outlines eight articulatory parameters as the basis for judging sonic modality: pitch extent, durational variety, dynamic range, perspectival depth, fluctuation range, friction range, absorption range and degree of directionality. As briefly touched on above, the modality of sound requires not only that its coding orientations be defined but that a distinction be made between representative and presentative modalities under each coding orientation. Van Leeuwen summarizes: “Presentation occurs when sounds are used to enact meanings in the here and now, rather than to refer to something that is not present in the here and now of the communicative event” (1999: 210). Representation, conversely, occurs when sounds are used in the latter manner; sound can be used to represent our environment and the actions and interactions of people, imitating sounds or distilling key qualities from them (1999: 93). Whether used for presentation or representation, Van Leeuwen maintains that sounds are ultimately dynamic, not things, nor can they be used to represent things. Being temporally tied, “sounds are actions and can only represent the actions of people, places and things” (1999: 93).
Abstract-sensory coding orientation combines elements from both abstract and sensory coding orientations, as it is rare for sound production to be completely one or the other. This is because, Van Leeuwen argues, even though music is the most abstract form of sound, it is almost impossible to be dispassionate and unemotional in the medium of sound (1999: 177). While the structure of music can be rationalized, it does not mean that its pitch and/or duration range is reduced. Both may be “exceptionally wide, and hence exceptionally charged with emotion”, which “makes it possible for music to be both abstract and sensory, and to aestheticize the sensory at the same time as it emotionalizes the abstract” (1999: 178). For representation, high abstract-sensory modality requires simultaneous presence of abstract representation and emotive effect. In case of presentation, high modality can be found in ritualized and emotionally restrained forms of interaction (1999: 181).

Naturalistic coding orientation, in the case of both presentation and representation, requires a wide range of parameters of articulation for high modality. Representation bases naturalistic modality on a criterion of verisimilitude: “the more a representation is felt to sound like ‘what one might hear if present at the represented event’, the higher its naturalistic modality” (1999: 181). Presentation bases naturalistic modality on ‘everydayness’ and ‘normality’: “the more a sound event is felt to be neither ritualized and formalized, nor dramatized and infused with emotion, the higher its naturalistic modality” (1999: 182).

Sensory coding orientation, finally, presents a type of modality where presentation and representation merge: its truth criterion lies in the degree to which a sound event is felt to have an emotive impact (1999: 182). As Van Leeuwen puts it, “The aim of the seductive advertisement is not to represent seduction, but to seduce. The aim of the horror film is not to represent horror, but to horrify” (1999: 179). Here high modality is typically achieved by amplifying articulatory sound parameters.

2.4. A note on types of modality

This study is mostly concerned with linguistic modality as semioticians see it. As
discussed above, Van Leeuwen points out that “modality expresses not so much the actual truth, but the truth as seen by the speaker: speakers or writers assign modality to their assertions to express as how true they would like these representations to be taken” (1999: 156). Similarly, different video games have different priorities in presenting objects within their game worlds to the player, with some objects being assigned higher modality than others as dictated by the needs of the game and the genre. This principle applies to both graphical and sonic objects within video game worlds, and this study will focus on the latter type. I will not be using the term ‘modality’ to refer to the different possible channels via which signs may be transmitted, even though most if not all of the 'sonic modalities' under scrutiny are indeed multimodal (simultaneously both visual and aural) in nature.

As game sound is never completely separate from events taking place on the screen, it is challenging if not impossible to assess the significance of a given sound event for the total audiovisual experience of immersion without simultaneously and thoroughly assessing the influence visual input exerts at any given moment. There would be ample material for detailed analysis in just a few frames captured from screen and combined with sound, much more so in a few minutes of video, let alone a complete play-through spanning perhaps hundreds of hours of gameplay. Bearing the above considerations in mind, there is nonetheless merit in analyzing subjective reactions to sound events on a broader scale, as part of fully-fledged soundscapes in comprehensive game audio, while acknowledging the visual context on a more general level. This is the approach I will adopt in the present study.

2.5. Affect and expectation

A widely-cited work among studies on the topic of music and emotion is Leonard B. Meyer’s treatise, Emotion and Meaning in Music, published in 1956. Meyer bases his musical analysis on a general psychological theory of emotions, contending that “though the stimulus situation may vary indefinitely, the conditions which evoke affect are general and are, therefore, applicable to music” (1956: 22). In this study, I will take this idea one step further and presume that these conditions are indeed
applicable to all sound in general and artificial sound in particular. I define the term 'artificial sound' in broad terms and include under its umbrella all possible sonic events in a constructed soundscape, be it a song, for instance, or complete game audio. The central difference between natural sound and artificial sound is the environment or the context it occurs in and therefore the reactions it can be expected to evoke in a listener. A natural soundscape is often by necessity filtered due to an overabundance of sonic stimuli, as pointed out earlier by Schafer. An artificial soundscape, on the other hand, is made to be listened to; if not actively, then at least passively.

Meyer’s analysis is primarily concerned with artificial soundscapes, or to be more precise, the sonic construct called music. Purposeful listening typically becomes a matter of passive concentration, almost completely negating the need for filtering out unwanted sound and sensitizing the listener for affective experience. As Meyer points out, the “listener brings to the act of perception definite beliefs in the affective power of music. Even before the first sound is heard, these beliefs activate dispositions to respond in an emotional way, bringing expectant ideo-motors into play” (1956: 11).

Central in the psychological theory of emotions is the idea that “emotion or affect is aroused when a tendency to respond is arrested or inhibited” (1956: 14). A tendency is “a pattern reaction [...] which, once brought into play as part of the response to a given stimulus, follow[s] a previously ordered course, unless inhibited or blocked in some way” (1956: 24). Tendencies may be natural or learned, and they may be conscious or unconscious. “If the pattern reaction runs its normal course to completion, then the whole process may be completely unconscious. [...] The tendency to respond becomes conscious where inhibition of some sort is present, [...] Such conscious and self-conscious tendencies are often thought of and referred to as ‘expectations’” (1956: 24). In a microlevel analysis, Meyer describes how a given musical gesture, say a few notes, will produce a tendency in the listener towards hearing a typical continuation for it in order for the notes to reach completion, for example, in form of a popular melody. He points out that expectation can be
understood in a much broader sense as well, or it can refer to a much broader object, perhaps the structure of a whole musical composition. What all types of expectation have in common is that the exact type of tendency is triggered and defined by the listener’s memory and enculturation (1956: 62), meaning that people of different backgrounds will likely experience different tendencies or different resolutions for same tendencies.

Although Meyer acknowledges a consensus of sorts that all types of music are seen to by definition have the ability to provoke at least some kind of emotional response in the listener (1956: 6), it has been a matter of some debate “whether music can designate, depict, or otherwise communicate referential concepts, images, experiences, and emotional states” (1956: 32-33). A central difficulty has been that discussions of the meaning of music have also been muddled by the failure to state explicitly what musical stimuli indicate or point to (1956: 34-35). Meyer attributes this problem in significant part to the difficulty of communicating and verbalizing affective experience, taking into account the myriad of cultural backgrounds against which these experiences take place. Music also has the ability to trigger memories, which in turn may give rise to affective experience. This may result in mistakenly attributing to the music affective experience triggered by unrelated reveries (1956: 9).

Due to the inherently subjective nature of affective experience, it is difficult to objectively measure anything in and of the emotive realm. Meyer, however, rejects “the traditional dichotomy between reason and emotion” and maintains that “there is no diametric opposition [...] between the affective and the intellectual responses made to music. Though they are psychologically differentiated as responses, both depend upon the same perspective processes [...] and the same musical processes give rise to and shape both types of experience” (1956: 39-40).

Juslin (2009: 131-132) defines affect as follows:

Emotions belong to the field of affect. [...] affect is regarded as an umbrella
term, which includes various affective phenomena such as preference, emotion, and mood. [...] Many researchers assume that people are always in some affective state. When the states are intense and involve salient stimuli, we tend to call them ‘emotions’, whereas when the same states are less intense, and their causes are not immediately apparent, we tend to call them ‘moods’.

Meyer also makes a point of distinguishing between emotion and mood, maintaining that many studies claiming to analyze emotion in music have actually more to do with mood and association than actual affective experience (1956: 7). While Juslin sees mood as another, less pronounced type of affective experience, Meyer places mood outside of the realm of affect. Both Meyer’s and Juslin’s studies agree that mood can be understood as a continuous, relatively stable state, while emotion is a more evanescent, more pronounced and continually changing experience. Meyer focuses on the latter type of experience in his study, while I, following the above definition set by Juslin, will be discussing elements of both under the umbrella term of affect in this study.

2.6. Distance and perspective

I will lay out some of the key similarities and differences in sound design between games of different genres from the viewpoint of distance and perspective before moving on to analyze my corpus in the following chapter.

First of all, the soundscapes of all video games are by default hi-fi, when in fact a lo-fi soundscape would be more representative of the real world. Video games contain no unwanted sounds; no noise whatsoever. Every sound event present in a game has been included for a purpose and should therefore at least be heard if not always actively listened to by the player. However, moments of silence are almost non-existent, since background noise from the real world tends to act as a forceful reminder of its presence and subtract from immersion in the game. This concern is in part mitigated by subjecting the player to a constant barrage of sound with dynamics carefully designed so that, instead of tiring the player's ears, it actually makes the player wish for prolonged exposure. The barrage of sound acts both as stimulant and as audioanalgesia against sounds of the outside world.
As noted earlier, games of different genres have different requirements for both graphics and audio in order to most effectively communicate features of objects within their respective game worlds to the player. A way of differentiating between games of different genres or different types of games within the same genre can often be found in the analysis of their intended points or loci of identification (my term). A locus of identification denotes both the point of view a game positions its player in and the relationship the player can be expected to establish with the game world, and a video game soundscape is typically designed around one such given locus.

Van Leeuwen (1999: 27) discusses the idea of social distance as regards the use of the human voice. Much like Schafer does, Van Leeuwen sees the invent of recording and amplification as having resulted in the sound of voice becoming an independent semiotic system, “able to create imaginary social relations between what is presented or represented by a sound and the listener”. Acknowledging a continuum rather than a strict division, he outlines five categories for social distance: intimate, personal, informal, formal and public. In speech, intimate distance is realized by whispers or maximally soft voices; personal distance by a soft, relaxed voice at low pitch and volume; informal distance by a more businesslike, full voice; formal distance by an overloud, higher and tenser projected voice; and public distance by the maximally loud sound, shouting at the top of one's voice. Van Leeuwen makes the important observation that these relations can also extend to places and things.

Strategy games typically cast the player into the role of a god-like entity (sometimes literally) which, while controlling almost everything that happens on a larger scale, may be less concerned with the minutiae of the dreary existences of those underneath it. As the locus of identification is far above the game world, the player will not expect to hear sounds that would imply any kind of special closeness with objects on the screen. This perspective, using Van Leeuwen's terms, usually applies public or formal distance in the use of most sound effects in the game. Zooming in is possible to a degree in some strategy games, but this seldom has more than a marginal effect on the soundscape in terms of renegotiating the player's position in relation to the
game world. A top-down isometric (i.e. birds' eye view) strategy game – let us say, *Civilization V* by Firaxis Games – usually features a pervasive ambient music soundtrack as its field, but stylistically the music tends towards the low-key end of the dynamic spectrum. The style (if not always the volume) of the music can, however, at times approach both the bombastic and the fairly intimate, and it makes use of most distance categories ranging from public or formal to personal. The background music may change, for instance, corresponding to an era or a culture the gameplay situation involves or varying gameplay states – for example, switching musical styles between states of war and peace – but as a whole it will usually remain inconspicuous enough to let the player fully concentrate on building and devising stratagems.

Strategy games typically contain an element of micromanaging cities, buildings and/or mobile units (soldiers, builders etc.), and audio feedback is key in their effective control – especially in real-time strategy games in which accuracy and speed are paramount for successful play. Units respond in a manner unique to the unit type when selected and given commands to, making the player less reliant on visual confirmation. Unit feedback is typically a brief, general verbal acknowledgement or a sound effect that is clearly discernible over all the other elements in the game’s soundscape at that moment and unambiguously identifiable with the unit in question, and the social distance of these responses tend toward the informal or even personal.

A fast-paced arcade-like driving game (for instance the *Flatout* franchise by Bugbear Entertainment) usually has its preferred locus of identification a few metres behind and above the vehicle, while simulator-type driving games generally prefer the first-person perspective inside the vehicle. The first locus lies within an informal or formal social distance, while the second would clearly be intimate. Some of the first and foremost priorities in arcade-like driving games are convincing if not downright awe-inspiring engine and tyre friction sounds. Not only are revving engines and screeching tyres powerful emotive signals, but directional and dynamic effects enable the player to better assess distances between vehicles (decent rear-view
mirrors being something of a novelty in most games) and are therefore of crucial importance for in-game success and the gaming experience as a whole. Also on the foreground is music, typically of rock or another upbeat genre, which serves to create and maintain a level of excitement, but environmental sounds are usually kept to a bare minimum with the notable exception of crashes and collisions. There are, in other words, fairly few keynote sounds in the soundscape of the stereotypical arcade-like driving game while figure and field (the latter in form of music) are prominent. Simulators, on the other hand, tend to have less ambient music outside of game menus and are therefore more concerned with realistic environmental sounds – naturally to the extent they can be expected to penetrate the vehicle hull. The car sounds, both alone and combined with music, have a strong masking effect in both types of games – if there are other elements in the soundscape, they cannot at most times be clearly heard over the din. It should be noted that most driving games include the option to switch between camera modes. If the point of view is outside the car, the greater masking presence is usually music, and if it is inside, the body of the car itself filters out or muffles most environmental sounds.

First-person shooter (abbreviated as FPS) games aim at creating the illusion of actually being the character or at least of observing game events through the character’s eyes. Of all video game genres, FPS game audio tends to be the most concerned with making the game environment as acoustically ‘real’ as possible. FPS sound design involves taking into account directionality, reverberation, contact with different materials and surfaces, and many other factors in a much more detailed manner than is strictly necessary in games of other genres.

The focus of FPS game audio tends to be on figure, ground and keynote sounds, while the field may be realized in the form of background music or simply more ground and keynote sounds that together form a general ambience. FPS games, as the moniker might suggest, have their preferred locus of identification inside the player character’s (PC’s) head, and as such the player should ideally see and hear everything the character sees and hears. As the player’s attention is directed via the character’s active gaze which can only cover a small portion of the game
environment at a time, the player effectively becomes reliant on the only other available sense, hearing, to impart information concerning areas in the game environment the character has no visual access to. Especially in more action-oriented titles, directional sonic cues are significant in alerting the player to danger and enabling them to react. FPS game audio makes use of the broadest range of social distances across game genres, all the way from intimate to public. This makes FPS games uniquely equipped to create and manipulate soundscapes where the players' expectations can be toyed with and new, exciting relationships between objects in the game world can be established. While many FPS games give the player the option to zoom out (usually behind and above) from the character, this change of perspective does not always coincide with similar changes in the soundscape, sometimes making the transition feel somewhat artificial.
3. Corpus and analysis

Fallout 3: New Vegas (henceforth abbreviated as F3:NV) is the sixth instalment in a series of post-apocalyptic role-playing video games originally developed and published by Interplay Entertainment, currently developed by Obsidian Entertainment and published by Bethesda Softworks. The style of the series is informed by the general 1950s' Cold War aesthetic and also likely inspired in significant part by Poul Anderson's novel *Vault of the Ages*, published in 1952. A basic premise in both the novel and the video game series is survival in a retro-futuristic environment blasted back to the Stone Age by a world-wide nuclear holocaust.

In terms of gameplay, the series started out as a partly real-time and partly turn-based strategy/adventure/role-playing game observed from an isometric perspective, but the two latest titles (Fallout 3 and F3:NV) have crossed over into the realm of the first-person shooter. I have chosen F3:NV as my corpus in this study because it contains elements from both the FPS and the role-playing game (RPG) genre. A certain intimacy, realism and high fidelity are usually required of FPS audio particularly where sound effects are concerned, while RPGs typically feature large, sweeping and emotive orchestrations as well as low-profile ambient music. F3:NV makes the attempt to combine the best of both worlds, so to speak, and it should provide me with enough source material to make arguments applicable to both of the above genres.

I will start the treatment of my corpus by first examining the roles and functions different types of sound events have in the game using Schafer's idea of perspective and distance as an aid. I will then examine the sounds further according to their modality, coding orientations and articulatory parameters as outlined in Van Leeuwen's theory, taking into account the probable circumstances the sound events were designed to be used in. Finally, I will discuss how the reception of a sound event is always affected by what has preceded it and what is likely to follow, basing my thoughts on Meyer's ideas of expectation and tendency.
3.1. Soundscape of the wasteland

I will begin my analysis by categorizing sound events in the game under music, sound effects and the area where the boundaries between the two blur. Borrowing terminology from film studies, I will split these categories further by diegesis into diegetic sounds (sounds that clearly have their source of origin somewhere in the game world) and non-diegetic sounds (sounds that clearly do not). As occurs with music and sound effects, instances of borderline diegesis can also be found. It should be noted that the concept of diegesis is in many ways very similar to Van Leeuwen's distinction between presentative (sounds used to enact meanings in the here and now) and representative (sounds used to refer to something not present in the here and now) modalities, but these terms are not, however, interchangeable. Diegetic music, for instance, may be both presentative and representative at the same time, as I will discuss in the next section. Using Schafer's terminology, I will discuss the function and role of each sound event under scrutiny in the soundscape. Perhaps risking stating the obvious, it should also be noted that the player will have some control over the different elements of audio playback in F3:NV's audio menu mixer, and its adjustment will affect the gaming experience as a whole by emphasizing some aspects of the soundscape at the expense of others.

Most sound events in the game world have a set range, be their sources stationary or mobile. Some sound events occur evenly across an area, while sources of others can be located in space with fair precision. Most ambient sound effects and music are location-specific, although the size and range of a ‘location’ may range from anywhere between a few metres to hundreds of metres. A small diegetic sound effect location might, for instance, be a gurgling toxic pond with a diameter of just a few metres, while a larger one could theoretically be a waterfall (although a dustfall might be a more realistic encounter in the parched wastes) or an entire mountain range. Diegetic sound locations can be likened to ripple effects in a pond: the closer the listener moves to where the sound location is centered, the higher its volume and intensity. Non-diegetic sound locations, on the other hand, tend to be of fairly uniform, lower volume and of larger diameter. Transitions between sound locations...
are smooth of necessity, as the player always has available the option to change
direction or walk away at any point. As Obsidian Entertainment’s audio director
Scott Lawlor (2010: 2) elaborates, sound locations can be designed dynamically
using multiple layers: “As the player starts to see a house on the horizon, the first
layer (of three) starts to play. The player hears the tension change. As he nears the
house, a second layer comes in, and once he is in the center of town, the whole music
track plays.”

Inon Zur, the composer of the original music in F3:NV, retained the services of The
Lyris Quartet for the majority of the game soundtrack in addition to including some
of the well-known pieces from earlier instalments of the game series, complementing
otherwise synthesizer-driven audio with a more traditional, gritty and intimate sonic
element. The string quartet – consisting of cello, viola and two violins – is
undeniably a central distinguishing factor in the soundscape of the wasteland, and, as
I will point out in the next two sections, it is responsible for much of the affective
experience produced by the game’s soundscape as a whole.

3.1.1. Sound effects

The most immediate and urgent subgroup of sound effects are diegetic sound effects.
Many of the most prominent diegetic sound effects have the role of signals within
the soundscape, and they include sound events such as footsteps, gunshots and
spoken dialogue, i.e. voice-acting.

As the player character moves everywhere in the Mojave wasteland by foot, footstep
sounds rank among the most repetitive sound effects in the game and are therefore
surprisingly important for the soundscape as a whole. Walking, running, sneaking
and jumping each have their respective sets of footstep sound effects in F3:NV, and
these footstep sounds are modified by the game engine in real time to adapt to the
type of surface the character is moving on. In order to minimize the numbing effect
of repetition, the game engine makes small, randomized dynamic changes to the
pitch and intensity of the footstep sounds. The variations are not very drastic, and
after short amount of gameplay the player is liable to become accustomed to constantly hearing the footstep signals, which would at this time partially recede into ground as a result. The shuffling and rustling of clothes during character movement and the drawing and holstering of weapons are likewise commonplace diegetic sound events. The rustling of clothes are linked to footstep sounds and could be characterized as something between signal and ground, whereas drawing and holstering have a function closer to that of signals.

Weapon sounds, another very common type of sound event in F3:NV most definitely acting in the role of signal in the soundscape, tend to be emphatic and well-defined, often with a clear in-built hierarchy: the more efficient a gunpowder weapon is portrayed as in the game, the more likely it is to have a resounding effect when fired, usually in form of added bass frequencies and a longer tail of reverberation. Weapons capable of rapid fire, however, do defy this ‘rule of bass’, as their bursts also need to be well articulated, and many consecutive bass-heavy sound events have the effect of muddying up the lower end of the sound spectrum (and overlapping frequencies is a concern a sound designer has to be very conscious of on every level of sound design). Melee weapon swings follow the same principle, as larger weapons are usually more effective and increased air resistance is represented by a louder, lower ‘whoosh’. Energy weapons and hand-to-hand combat sounds follow the principle to a lesser degree.

Voice-acting in video games has been a growing trend in recent years, and it is nowadays viewed as an almost mandatory part of story-driven RPGs in particular. F3:NV is no exception to this rule, and every line spoken by an NPC (non-player character) is voiced. The game world of F3:NV is very large, and it features numerous NPCs with unique dialogue written for most of them. All instances of voice-acted sound events are initiated by the player directly (by activating an NPC) or indirectly (by the proximity of the PC to an NPC), and they all have the role of signal in the soundscape (sometimes even when not entirely appropriate: see voice-acting and modality in the next section).
Other diegetic (mostly, see below) sound effects include environmental signifiers, which often vary according to the time of day in the game. These type of sound events are not initiated actively by the player – they are instead heard by being in the right place at the right time. Daytime environmental diegetic sound effects in the wild of the Mojave wasteland include birdcalls and distant, echoing gunshots, and their dynamicity is likewise randomized by the game engine within specified keynote parameters. Wind can be heard howling in canyons and rustling through dry grass and dead foliage day and night. Night-time tends to otherwise be a bit more ominous: birdcalls turn to hoots and howls, legions of crickets invade the soundscape, and every once in a while the ambience is punctuated by shutters clanking in the wind, tin cans or similar metallic objects dropping to the ground and by distressed, very distant human vocalizations. Caves and caverns always feature a prominent diegetic wind effect echoing through the passages in the role of ground. All of these ambient effects, even if occasionally mildly startling, are of fairly low volume and should be categorized under the domain of ground and field.

Non-diegetic sound effects are also numerous in the game, and they are typically used to impart meta-level information to the player. These effects occur, for example, when the player character becomes stronger by gaining a level (a brief, triumphant tattoo is heard), when the character receives, completes or fails a quest (low thudding sound or a falling note), when the character gains any number of experience points (low thud), or when the character’s karma value changes (jingle for good, falling note for bad). These bits of supplied meta-level information typically require immediate attention from the player and should therefore be regarded as signals in the soundscape. There is also some non-diegetic voice-acting at the end of the game and both at the start and the end of each additional downloadable content module, likewise in the role of signal. This voice-acting takes the form of a third-person narrator who briefly describes the characteristics of a new area or recounts the most significant outcomes of the PC’s actions and the effects they have had on the Mojave wasteland.

Borderline diegesis can be found in a device called the Pip-Boy the PC wears on one
arm. The Pip-Boy is basically a wrist computer the player uses to interact with the character's inventory, maps, quest logs and other data. Although every sound event from beeps, button clicks and toggles to the four variations of the buzz the Pip-Boy emits as it is brought up and put away is believable and realistic within the game world's parameters, the game pauses every time the Pip-Boy is brought up. This, in effect, places the device almost but not quite outside of the game world and makes the sound events associated with it borderline diegetic by definition. As changing clothes and armour as well as equipping and unequipping weapons are done via the Pip-Boy alone, all of the sound events involved in these actions are clearly borderline diegetic. All of the sound events involved in operating the Pip-Boy inform of immediate changes made by the player and should therefore be regarded as signals.

The pounding, insistent heartbeat that can be heard once the player has become sufficiently wounded would be a prime example of a diegetic signal were it not for the fact that the heartbeat does not pause when the Pip-Boy is activated. One could argue that this fact places the sound event, like the Pip-Boy itself, partially outside of the game world, although the heartbeat does exist outside of Pip-Boy as well. I would probably still prefer to think of the heartbeat as being closer to diegetic than non-diegetic, and the effect does help in creating the illusion that the surrounding world does not stand still while the PC is rummaging through a backpack looking for a cure for a gunshot wound when it in fact does. It should be noted that environmental effects do not pause either when the game is paused by bringing up the Pip-Boy, and thus they enforce the illusion of time not standing still. They cannot, however, be replaced by a different set of environmental signifiers without the PC physically moving or the game clock advancing, neither of which is possible while still in the Pip-Boy menu.

The Pip-Boy's many functions also include a 'VATS targeting system', which can essentially be used to pause the game and queue up actions during combat. When these actions resolve, they (and their associated sound events as signals) are presented in slow motion using cinematic camera angles. Slow motion naturally
implies slower movement and stretched sound, which is of lower frequency (pitch) and longer duration than its normal variant. Even though the time flow of the game world is presented differently, the sound events of actions resolving in VATS mode take place in the same space and time as the rest of the activity in the game world, and I therefore prefer to regard these sound effects as diegetic instead of borderline diegetic.

3.1.2. Music

Diegetic music in the Mojave wasteland comes in the form of pieces of popular music from the decades preceding and following the 1950s, invoking, on the one hand, the Cold War aesthetic and the nuclear scare of the time and a sense of technological optimism on the other. These musical pieces can be heard playing on several radio stations, most of which have coverage all across the wasteland. Radios can usually be found within urban centres, houses and also scattered outside in the wild, but one radio is also permanently strapped on the PC's wrist to be used at leisure. The Pip-Boy's radio can be activated at any time during gameplay, effectively overriding any other ambient music in the background until turned off. Again, listening to the radio while the game is paused in the Pip-Boy menu presents a temporal detachment, making the music heard borderline diegetic instead of diegetic. The radio programs additionally feature news of the wasteland, switching the radio's role between signal and ground. Diegetic music is also performed live at one of the casinos in the New Vegas Strip, arguably in the role of signal in the soundscape.

Non-diegetic music is present everywhere in the game in the role of ground/field, and it is used to convey a wide variety of moods, connotations, and spatial and contextual information to the player. A heavy reverb effect, for instance, can often be used to underline the barrenness and openness of the wasteland, while a lone violin playing minor-scale melodies might evoke feelings of melancholy and isolation appropriate for the setting. The typical ambient background music track takes the form of a set of fairly short sequences on separate layers mixed together by the game
engine, and the choice of the pool from which the sequences are drawn depends primarily on the physical location of the PC. A sequence might, for instance, be a falling and rising air raid siren used as a tonal element in the background of the mix, acting as a subtle reminder of the war and catastrophe that has befallen humankind, or it might be a stripped-down version of a theme to be accompanied and complemented by other similar sequences. There are many instances of specific music locations in the game, which are commonly used to signify the presence of a faction or a geographical feature. For instance, distant drum tattoos and solemn or bombastic brass instruments can be heard when the PC approaches a military encampment, while a casino on the Strip may feature anything from sultry jazz to classical music, depending on the style of the venue. Large sewers, vaults and other underground locations usually feature a set of forbidding ambient tracks that often allude back to earlier instalments of the game series, while caves and caverns sometimes have nothing but their accompanying diegetic, environmental sound effects.

Combat music is the only type of non-diegetic ambient that fills the roles of both signal and ground in the soundscape. As battle ensues, the uptempo and menacing combat music has a clear onset (a brief percussive intro), and when it finishes, a less intense outro plays. Although the music starts with a small delay and only after another denizen of the wasteland has become aggressive towards the PC or, if aggressive to begin with, notices the PC, the immediacy of the switch to combat music often acts as an important signal that alerts the player to imminent danger.

3.1.3. Blend of sound effects and music

Non-diegetic ambient music and diegetic background effects such as wind are often very close to one another both in texture and feel, and it could be argued that environmental sound effects may at times be regarded as being part of the background music. The argument works both ways: as the player subconsciously accepts some environmental signifiers as part of the background music, also some musical elements resembling, for instance, electrical crackles, hums and hisses of
heavy machinery as come closer to being accepted as diegetic effects. The player can almost imagine these sounds as being part of the natural soundscape outside in the futuristic wasteland, perhaps as the lingering results of atomic devastation. Inside vaults, sewers and other urban areas this is much less of a stretch of imagination, and it could be also argued that this fusion of music and effect could be seen as another form of borderline diegesis.

3.2. Modality and affect in F3:NV

In this section, I will examine possible modality configurations in the soundscape detailed in the previous section and their potential functions in inducing affect in the player-listener. Using previously discussed sound events as my examples, I will first determine which coding orientations should ideally apply to the sound categories examined above in F3:NV in general, depending strongly on context and their role in the soundscape. I will then identify which articulatory parameters are specifically valid for these modality configurations and how they are realized in the sound events under scrutiny. Different types of sound events have different specific gameplay functions, but they have also been designed to combine in a manner that produces a deeper experience of immersion for the player. The ways in which they contribute to this common goal depends on their preferred modality and coding orientations.

I will assess the soundscape of F3:NV in terms of naturalistic, sensory, abstract-sensory and technological coding orientations. Sound effects – presentative and representative – should ideally resemble their real-life variants enough to be convincing, but they should also make an emotive impact especially when filling the role of signal in the soundscape. Hence, the typical modality preferences designated to them (regardless of diegesis) are naturalistic coding orientation and, to slightly lesser degree, sensory coding orientation. In ambient music, on the other hand, a representative abstract-sensory coding orientation is by default preferred. Music can be rationalized in form (abstract), but it almost inevitably also evokes some kind of emotive response (sensory). The background ambient music in F3:NV is usually acting as ground or field, making it important and well-placed for creating and
maintaining mood, but it does have some recurring elements that occasionally rise to the prominence of signal in the soundscape, sometimes evoking identifiable and specific emotive response. I will return to technological coding orientation later in this section.

The definition of high-modality naturalistic coding orientation has a wide range of articulatory parameters and is based on criterion of verisimilitude in case of representation and on criterion of everydayness or 'normality' in case of presentation, while sensory coding orientation amplifies some or most articulatory parameters for maximal modality and effectiveness in producing affective experience in both cases. Abstract-sensory coding orientation is seen as being of high modality when, in case of representation, abstraction and emotive effect are both present, while presentation focuses on ritualized and restrained forms of interaction (Van Leeuwen 1999: 181-82).

3.2.1. Diegetic sound effects

Diegetic sound effects in F3:NV – presentative in particular – tend towards mimicking realistic 3D environments especially in terms of absorption (space, reverb, echoes) and directionality (panning and volume adjustments of multiple channels starting from stereo), as would most FPS games. As footsteps, weapon sounds and voice-acting are presentative sound effects acting in the role of signals, their primary modality preference would be naturalistic coding orientation based on the criterion of normalcy with just a hint of sensory coding for that slightly more-than-real emotive effect.

F3:NV features several sets of cleverly designed footsteps that correspond to different surfaces and textures (sand, stone, gravel, metal etc.) and different types of movement (walking, sneaking, jumping, running) on said surfaces. The articulatory parameters cueing naturalistic coding are all geared towards a fair level of amplification, which added sensory coding will slightly increase. Footstep sounds have a fair degree of *durational variation* in both different modes of movement and
on different surfaces. A single running footstep sound on gravel, for instance, is a sharp thud with a very short crunchy tail. A walking footstep on the same surface is slower with a prolonged tail, while in sneak mode both of these sound events exhibit same behaviour, although muffled. The sneak mode thus increases dynamic range of the footstep sounds, while falling or landing from a jump on any surface increases it in the other direction by producing an effect noticeably louder than walking or running. Still using movement on a gravelly surface as my example, the footstep sounds feature a degree of perspectival depth, such as can be expected. On the other hand, they produce no discernible reverb or echo effects, and as such the absorption range of these sound events is rather small. Directionality is static as can be expected, as the distance and relative position between the PC’s ears and feet never changes. Degrees of fluctuation is not really a relevant parameter here as it is mostly associated with vibrato in the human voice in Van Leeuwen's study, but the small pebbles crunching underfoot can certainly be said to have enough degrees of friction for maximal naturalistic modality. Pitch range, however, is where the footstep sounds fall a bit short. While the sound events are convincing enough when analyzed separately, the issue of repetition has been tackled by means of using pitch randomization that can at times sound slightly unnatural.

Weapon sounds, a crucial element in almost any action game, are emphatic and meticulously designed. Gunpowder weapons (except for the tiniest of pistols) sound powerful and effective, and even fantastic weapons with no real-world counterparts (such as plasma rifles, gatling lasers and other energy weapons) sound exactly like one might imagine them to sound, informed by the retro-futuristic aesthetic. There is considerable durational variation between the different weapon sounds, which as a whole have a fairly broad dynamic range and pitch range that are commonly manipulated to indicate how powerful a weapon should be seen as. Projectile weapons feature some perspectival depth, as gunfire and explosion sounds do behave differently depending on the distance between the PC and the sound source. As noted earlier, gunshots also use a number of different reverbs, often as indicators of power, finding a use for a fairly wide absorption range in addition to enhancing the perspectival depth of the sound event. Directionality is well present, and ricochets
and explosions create their own, pinpointable sound events. Energy weapons perhaps have a little fewer *degrees of friction* than other weapon types do, but their futuristic nature lends itself well to a more synthesized type of sound, which is still quite gritty.

Voice-acting tends to be crystal clear, almost to the point of surreality. Although dialogue is consistently performed, recorded and mixed very well and normally of very high naturalistic modality, it often assumes standard speaking range even when the PC should clearly not be able to hear the speaker well (or at all). What is much more jarring, being underwater does not hinder the PC's NPC companions from commenting on various things in the slightest, and neither the voice-acting nor any of its associated effects reflect the state of being immersed in water. In terms of articulatory parameters, *pitch range* of voice-acting is naturally very wide, as is *durational variation* and *dynamic range*, and the timbre of a human voice naturally has many, many *degrees of friction*. *Directionality* is as realistic as with any pinpointed sound location, but the problems mentioned above arise in the low amplification of parameters *perspectival depth* and *absorption range*. *Degrees of fluctuation* is still not a very relevant parameter, as very few human voices in the game do any actual singing.

Environmental signifiers usually act in the role of ground or field, and as presentative sound events they use the same basic set of modality preferences as the effects discussed above. The function of ground, however, arguably allows for some more leeway in terms of coding orientations: while ground and field are not designed to grab the player's immediate attention, they can and should subtly act both as a reminder of the PC's physical location and as a vehicle for affective connotation. Here, naturalistic coding orientation can perhaps be followed a bit less strictly, while sensory coding orientation can be afforded a somewhat larger margin of operation.

The number of different diegetic environmental effects in the game is very large, and they are much more variable than any of the other effect types previously discussed. Several environmental effects are typically playing on different layers at any given moment in the game, and they make broad use of all the articulatory parameters.
3.2.2. Non-diegetic sound effects

Non-diegetic sound effects typically act in the role of representative signals, and they are primarily used to impart meta-level information to the player relating to the PC's current situation, health, objectives, et cetera. These signals are much less concerned with naturalistic coding orientation than sensory coding orientation, and as such they are geared more towards coaxing an affective reaction from the player than diegetic sound effects are. As the PC gains a level (a very happy occurrence), the game engine plays a brief, triumphant drum tattoo. This sound event has obvious military connotations and fittingly so, since it is quite common that the player will increase one of the PC's many available weapon skills during level-up. The effect is uplifting, yet solemn. When the PC gains experience points (an increasing number of which is required for every consecutive level-up), the game engine plays a low, thudding sound with a long reverb and a single delayed echo. This sound event feels momentous and important even if it does not have any particularly strong connotation. Completing or failing a quest results in a similarly prominent thud, albeit with a more voiced, rising tail. When the character's karma value changes for the better, a bright jingle is heard, while a change towards the dark side is characterized by a falling, reverberating note, evoking and enforcing a feeling of disappointment.

Non-diegetic voice-acting follows mainly the same parameters as diegetic voice-acting does, but the lack of perspective and absorption are not really an issue as the narrator is not present in the game world.

3.2.3. Borderline diegetic sound effects

The background hum the Pip-Boy makes when brought up could be thought of as the only instance of a borderline diegetic sound effect in the role of ground. Otherwise, all effects produced by the Pip-Boy are signals responding to the player-character manipulating the wrist computer. These effects are both representative and presentative. For instance, when the character dons an armour, a very brief sound
event that resembles handling said material plays. When the character eats food, drinks liquids, equips weapons or handles other equipment, the sound events played have similar real-world resemblance, even if they are impossibly short. The sound events are, in effect, snippets standing for much longer processes. Borderline diegetic sound effects are more concerned with naturalistic coding orientation than non-diegetic effects are, but less so than diegetic sound effects. Affective experience is less important in the Pip-Boy than relaying unambiguous information about an action taking place, and so sensory coding orientation is not a first priority.

Although Van Leeuwen did not include technological coding orientation in his list of modality configurations for sound, I believe this orientation may sometimes be viable in the context of video game soundscapes. As discussed above, sound events are commonly included in video games for the purpose of informing the player of meta-level changes in the game environment and the player character's current situation or status. These meta-level changes typically have a direct bearing on gameplay. An example in F3:NV could be the thudding heartbeat that can be heard once the PC is sufficiently low on health, a borderline diegetic signal. The lower the character's life value drops, the more rapid and insistent the thudding becomes. The quickening, gradually almost deafening heartbeat is by no means representative of similar real-world cardiovascular stress – its tempo, volume and texture are metronomically constant unless the life bar lowers further across a predetermined boundary, and so its naturalistic modality is weakened by the low amplification of parameters dynamic range and durational variation. The effect, however, does serve to provide the player with fairly accurate information that has direct relevance for the PC's survival, quite possibly in a hectic situation where the player will have no time to actively look at numerical gauges or graphic bars on the screen. This auditive information guides the player's actions indirectly but efficiently towards more effective gameplay, and the sound event's modality could therefore be regarded as high in terms of technological coding orientation.
3.2.4. Diegetic music

As mentioned earlier in this study, diegetic music in F3:NV can be both presentative and representative, occurring typically in the function of ground or field. The primary sources of diegetic music in the game are several radio channels, broadcast across the wastes, that can be listened to at any time using the Pip-Boy. Music is performed live as well at The Tops casino. While the context of a musical performance would seem to imply that it should have the role of signal at the venue, it is mixed similarly to most other music that occur as ground.

Part of the 1940s-60s imagosphere, the songs used in F3:NV are classics that have both unintended stylistic and lyrical relevance for the post-apocalyptic setting. Many of the songs have added ironic and humorous undertones when reinterpreted in the bleak context of the game, and they can be thought of as being representative of the Cold War and the communist threat, a certain innocence and idealism of the era, technological optimism and the early science fiction aesthetic. Music in general follows the definition of abstract-sensory modality, and its articulatory parameters can be both reduced and amplified as befits the piece. The diegetic radio music, however, is primarily presentative and ritualized, and it follows the spirit and values set for the game world, the game series and the genre. Being restrained in this way, most articulatory parameters are reduced. It is hence understood that the songs are both culture-specific items and musical progressions.

3.2.5. Non-diegetic and borderline music

Although non-diegetic music in F3:NV follows the principles of abstract-sensory coding orientation much like diegetic music does, it is in many essential ways different from the radio music described above. Instead of being ritualized and presentative, the non-diegetic background ambient music is representative and adaptive, leaning towards the abstract side of the modality configuration in the sense that it is typically better equipped to represent underlying “essences” and common truths than diegetic music.
Ambient 'drone' music tends to be minimalist, and it can easily integrate many kinds of sound events either as they originally occur, as they could occur or as they occur but in a modified form. I will use the air raid siren song passingly referred to above as an example. The song, “Brotherhood of Steel” by Mark Morgan (also used in some of the previous instalments of the game series), is very calm and serene despite its military associations, evoking the sense of a distant memory rather than the present. The piece has no real discernible tempo, but the song has fluctuating dynamics that pace its progress. The rising and falling air raid siren used as a tonal element and the drum tattoos towards the end can easily be interpreted as symbols of war, screeching metal brings to mind heavy industry, while crackles, buzzes, hums and radio squelches have connotations to electricity and electronics. Connotations are instrumental in creating and maintaining mood, and, considering the sound events' role as ground/field in the soundscape, they often work subconsciously (sometimes the elements can be clearly heard from underneath all other simultaneous sound events in the soundscape, sometimes not as easily).

Ambient music of this type typically makes use of a very wide pitch range, although certain drone frequencies tend to be emphasized over other frequencies in the piece. Other frequencies represented are usually interjections in the form of shorter, more pronounced sound events. There is a lot of durational variation, as the steady, slowly evolving drone is periodically punctuated and decorated by these shorter sound events. Dynamic range is quite wide, but not so wide as to compromise the audio piece's role as ground/field in the soundscape. Perspectival depth is very wide, and the piece paints an almost surrealistic sonic picture with perceived distances that cannot readily be measured. There are some degrees of fluctuation, but this fluctuation is very slow and tied to changes in dynamic range – in effect, these two parameters merge. Shorter sound events, especially percussion and some of the metallic sounds occurring in the background, have more degrees of friction, but as a whole the sounds used in the piece tend to be more clean than dirty, i.e. more synthesized than natural. Absorption range is wide as well, and the heavy use of reverb and echoes lend to the piece a degree of etherealness and an impression of a vast surrounding space. There are some degrees of directionality to be found in the
way some elements are panned and the ways others echo away into the distance with a channel-alternating delay effect.

Slowly fluctuating song dynamics and almost no discernible tempo are a pervasive feature of other ambient songs as well, for instance the multi-tiered song “Wasteland” by Inon Zur. The piece is adaptive, playing in layers and fading in and out as the PC's location or situation in the wasteland changes, and in this way the player has indirect control over the soundscape in the game (in addition to direct control afforded by the mixer under the game audio settings menu).

“Wasteland” (as well as many other similar pieces in F3:NV) has a strong recurring element that stands out in the soundscape, it could be argued, in the role of a signal. The string quartet mentioned briefly in the beginning of the previous section are not the only clearly identifiable acoustic instruments used in the soundtrack of the game (the piano is quite prevalent, as is the guitar), but their parts are composed and mixed so as to become a distinguishing factor or a soundmark of the Mojave wasteland.

The occasional shaker-effect of the piece is reminiscent of a rattlesnake, and there are wind-like effects as well. The steel guitar and percussion sticks fit the wasteland cowboy aesthetic perfectly. Pitch range is wide, as is durational variation. Parameters perspectival depth and absorption range are skilfully amplified to create an impression of an impossible mix of wide and open spaces and the sound events that occur therein. A degree of directionality is primarily present in the form of panned wind effects and shakers, while the intimately gritty string quartet can certainly be said to bring quite a few degrees of friction into the mix. Dynamic range and degrees of fluctuation are the two parameters where this piece clearly differs from most other types of ambient music in the game. As noted earlier, composition and mixing have lifted some of the string elements into the domain of signals, typically in the form of one prominent solo instrument. The violin (or the viola or the cello, as the case may be) with its associations to isolation and melancholy, combined with a fitting reverb, is well placed and equipped to create a strong emotive response in the player. The dynamic range of the piece peaks with almost every instance of a soloing string instrument, and the degrees of fluctuation in the
strong vibrato the quartet players use are a prime example of how this articulatory parameter may be amplified to induce affect. Sliding notes would also fall under this parameter, and they are both beautifully executed and abundant in the piece.

3.3. Disconnect as agent of immersion

Determining the primary function of a sound event in a soundscape at any given moment is usually not enough to fully explain the onset of affect, however, let alone immersion. Sound events never occur alone, emotive response always happens in a context, and the experience of immersion requires a string of consistent emotive responses of sufficiently similar or at least compatible character. To be more precise, immersion can often be maintained simply by maintaining an appropriate mood, but increased engagement or deepened immersion nearly always requires an affective experience of some sort. Conversely, weakened immersion is usually due to an adverse emotive response or a failure to maintain a mood propitious for a particular type of gaming.

One of the biggest challenges in creating dynamic and consistent soundscapes for video games is trying to predict every kind of combination any two or more given sound events may occur in during gameplay. Sound events that have high modality by themselves may occur in a situation where the context of what is happening, what has happened just before or what is about to happen may be wildly inappropriate or unexpected. I have elected to call this discrepancy between a sound event and its context a ‘disconnect’, and while it may have either a positive (strengthening) or a negative (weakening) effect on immersion, it is typically unlikely to have no effect at all. Using Meyer’s terms, both types of disconnect deny the resolution of a tendency, producing affect of some kind.

There are some instances of weakening disconnect in F3:NV. It can be caused directly by low modality, as happened in the situation where an NPC would start easy chatting while diving deep underwater, but a weakening disconnect can occur even when modality, judged by itself, would be quite high. I witnessed an example of this
phenomenon right after a gunfight in the streets of Freeside, next to the New Vegas Strip. A barker at a street corner was ducked and cowering, and quite appropriately so – bullets had been flying here only a few seconds before. Less appropriate for the heated situation, however, was her scripted bit of dialogue, triggered when my character ran past just close enough: “Make a brief stop at the Atomic Wrangler, where the booze is cheaper, the tables more friendly, and the women are just like the booze!”

As discussed earlier, it is normative for first-person shooters and roleplaying games that the game music follow gameplay events as closely as possible. When nothing particularly exciting is happening, for instance while the player character is travelling down an empty road, one expects to hear low-key atmospheric ambient music appropriate for the game setting that is meant to be heard but not really actively listened to. On the other hand, when a confrontation is imminent, often the first cue the player gets of the rapidly changing in-game situation is the sudden outburst of music that tends to be drastically different from the default ambient: bombastic, threatening, nervous, upbeat and so on. During the shift from background ambient music to combat music, the function of the music elevates from ground to signal, its technological modality becomes higher, and all of this occurs suddenly. The player usually has no means to anticipate this abrupt change in the soundscape, which means that the tendency to hear a low-key continual similar to the background track that has been played just before is inhibited. The sudden elevation in dynamic range and other parameters produces a state of mild anxiety, increases the player's respiration and the adrenaline level in the blood. The result is a feeling of excitement and increased attentiveness. The sudden change in parameters is usually enough to cause this effect at the early stages of playing the game, but once the player becomes more accustomed to the rather common (albeit unexpected) sound event, having become conditioned to what follows (i.e. a violent conflict, its resolution and its consequences) is probably often as responsible for creating and maintaining a state of excitement as the sound event itself.

At one point during my exploration of the harsh and unforgiving wasteland, I
became aware of the presence of an approaching deathclaw in the pitch-black night of the Mojave desert. A deathclaw ranks among the most dangerous opponents the game has to offer, and for my relatively inexperienced character, meeting one so early in the game spelled certain doom. Surrounded by a cliff on one side, the geography prevented me from escaping, but the game engine had not yet been signalled that a battle was about to ensue (because the deathclaw had not spotted me yet).

The music I would have normally expected to hear in such a situation would have been the combat score or at least something very sinister. Instead, I heard ambient music punctuated by the eerie, hauntingly beautiful violin I had grown accustomed to hearing in peaceful, non-violent scenarios. Hearing the violin resulted in an experience perhaps best described as bordering sublime; a dreamlike sense of being there while not really being there, a sense of inevitability and not being able to influence the events that were about to take place. Due to my previously conditioned emotive response, the unlikely music in this particular context caused a disconnect between the intended message of the piece (such as there is) and my experience of it. Serene and melancholy music combined with the knowledge that somewhere out there in the night stalked a monster impossible to defeat and soon about to pounce made me experience a fatalistic sense of doom, dread and resignation – almost acceptance.

The disconnect did not weaken my immersion in the game, quite the opposite. Despite my character’s consequent (rather unfortunate) fate and the fact that I was forced to reload the game in order to continue playing, the unexpected and unintended element strengthened my emotional engagement in the game. Interestingly, due to this experience my appreciation of the violin score increased as well, and now whenever I hear the violin in question I am on some level reminded of this experience. Using Meyer’s terms, my tendency to hear the sound track change towards something sinister or bombastic was inhibited, but the resulting affect was intrigue and increased engagement.
As shown above, disconnect can occur via unexpected changes in sound event roles, in this case from ground to signal and with a positive effect. Van Leeuwen (1999: 19) draws a similar example from the movie *Piano*, where a scene culminates in one main character cutting off another's finger with an axe. Instead of diegetic sounds and effects, this horrific scene is dominated by non-diegetic music heard previously in situations associated with loss and longing. The victim of this violent act, Ada, responds by retreating to an inner, mental sanctum symbolized by a musical score the character has been heard playing before. As Van Leeuwen points out, Ada's inner world is effectively shown to be more present or relevant than the outer in the scene by the simple alteration of roles within the soundscape, and the viewer can identify with Ada's character and her situation better as a result.
4. Discussion and conclusion

During the course of this study, I have analyzed the soundscape in F3:NV first by seeking to determine the extent to which some of its most prominent sound events can be categorized under diegetic, non-diegetic and borderline diegetic music and sound effects, then by assessing the role (signal, ground and/or field) these sound events typically have in the soundscape, and finally by outlining some of the most likely modality configurations for these sound events. An effort has been made both to study generally some of the mechanisms underlying the experience of affective immersion and to figure out how sound events should be configured modally in order for the experience to have the best possible chance to come to fruition. A couple of questions arose during the analysis: if high sensory modality is grounded in emotive response that is also supposedly the prime requisite for any experience of affective immersion, why would sound effects still prefer a roughly equal mix of sensory and naturalistic coding orientations? Why would music likewise prefer the abstract-sensory orientation instead of concentrating on sensory coding?

It would seem that a balance has been struck between maintaining a believable fictional environment and providing an awe-inspiring soundscape. In the case of sound effects, the former is strengthened by high naturalistic modality and the latter by sensory modality respectively, and enabling and maintaining the player's suspension of disbelief is a delicate job not only in the realm of sound design. There should probably ideally be enough of a wow-factor present to keep the player's imagination aroused but not so much as to make the events portrayed on screen seem significantly less realistic as a result. In the case of music, the main feature that sets abstract-sensory coding orientation apart from and above sensory coding orientation is its ability to double its function as a vehicle for connotative thought process. Partly subconscious images and ideas of happier times long past and of nuclear devastation and desolation effectively lay the foundation for a more introspective, slower-paced play style that I find very suitable for the aesthetic of F3:NV in particular and the post-apocalyptic RPG subgenre in general.
In the present study, Van Leeuwen's theory of modality is generally understood as applicable to any sound event at any one given moment. One possible implication here is that the player-listener's enculturation, background and previous experience playing the same or another game do not really factor in cueing modality judgements. It would seem a gross oversight to neglect the importance of how previous experiences affect any current and future experiences of immersion, and the solution in this study has been to investigate this aspect of affective immersion through both Meyer's theory of tendency and expectation and my own idea of strengthening and weakening disconnect. It was found that surprise indeed breeds emotion; whenever a disconnect of either type would occur during gameplay, the experience of immersion would either weaken or be strengthened as a result. In the light of this study, it would seem highly unlikely that a disconnect could have little or no effect on immersion whatsoever.

The less prominent but nonetheless very important concept of mood propitious for continued gaming was not discussed in any particular depth, and the means of manipulating mood in sound design can likely be much more subtle than those used in inducing stronger emotive reactions. This would be a fascinating venue of further inquiry, as the current study concentrates on only a few rather rough-cut scenarios. It stands to reason that careful manipulation of mood could well make it possible – or at the very least easier – to consequently induce affective immersion in the player-listener in situations where such inducement would otherwise be difficult or impossible to achieve.
Bibliography


University Press.


Accessed online audio

<http://www.youtube.com/playlist?list=PL17D8535B1F2AE372> (All ambient music in F3:NV)

<http://www.youtube.com/watch?v=F0yTxs8GfoQ> (Brotherhood of Steel by Mark Morgan)

<http://www.g4tv.com/videos/49422/fallout-new-vegas-soundtrack-desert/>

<http://www.g4tv.com/videos/49424/fallout-new-vegas-new-california-republic-theme/>

<http://www.g4tv.com/videos/49423/fallout-new-vegas-soundtrack-mountains/>