8 Intermedial combinations

Mats Arvidson, Mikael Askander, Lea Wierød Borčak, Signe Kjær Jensen and Nafiseh Mousavi

Media products of all sorts form a complex web of different relationships. Media products involve transformations, integrations and combinations as well as transmedial aspects. When we look at media combinations in this chapter, all these different aspects are brought into play. Media combinations of different basic media types are always, literally, intermedial combinations that involve intermedial relations between different forms of communication.

This chapter deals with different kinds of combinations of technical, basic and qualified media types in comics, films, radio drama, songs/singing and music videos. Some of them are more obvious in this aspect. The music video, for instance, integrates sound, words and (moving) images. The pages of comics display text and image, and these two basic media types communicate differently in the semiotic modality. In radio dramas or songs, the combination aspect is not as visually apparent. Still, the soundwaves of a song or a radio drama firmly integrate several auditory media types.

With specific examples, we discuss how to understand the different intermedial aspects at play whenever different media types are brought together in a particular media product. Different perspectives are possible. Should one focus on the combination of different forms of meaning-making, or instead stress how deeply integrated these different media types are? Should one approach a song as the combination of different qualified media (poetry and music), or focus on the close integration of different auditory media types (words and organized sound)? When we approach media combinations with the four modalities, we can focus on both. When words, (moving) images, and organized sounds are brought together in media products, such as comics, songs and music videos, they form an integrated whole. We can focus on how different basic media types in the material and sensorial modality are firmly integrated. We can then explore how these integrations on pages, in soundwaves, on stages or in the studio enable an intricate combination of different forms of meaning-making that support and interact with each other in the spatiotemporal and semiotic modality.

First, we will take a look at how words and images on pages convey a graphic narrative in comics. Then we will highlight the importance of sound effects in the complex combination of moving images and auditory media types

DOI: 10.4324/9781003174288-10

in film. We then explore how different auditory basic and qualified media types together create a complex auditory narrative in radio drama, using the specific example of *The Unforgiven* (2018). We will also discuss how word and music combine on different levels in art and pop songs. The chapter will end with a few reflections on the audiovisual combinations of basic and qualified media types at work in music videos.

Words and images on the go: Intermedial meaning-making on the comics page

As 'a medium that communicates through images, words, and sequence' (Kukkonen 2013b, p. 4) or 'intermedial narratives based on words and images' (Rippl and Etter 2013, p. 191), 'comics' or 'graphic narratives' are qualified media that conventionally tell stories through an interaction between words and images on the page. The term comics is a general term that refers to a wide range of media such as comic strips, comic books and graphic novels, which differ regarding institutional conditions of publication and reception that are part of their qualified media aspects. The description 'graphic narrative' is being used more and more, especially in the research on comics, to be able to equally account for fiction and non-fiction comics and to emphasize the narrative work being done. Comics were initially tied to printed material and framed within other qualified media such as newspapers and magazines or published autonomously as books. Since the 1990s, the internet has been a popular canvas for the digitalized production and distribution of comics in different forms, widely known as webcomics. Professional or amateur comic artists have been able to create brief or lengthy graphic narratives in various hand-drawn or computer-generated formats and publish them in personal blogs or more institutionalized platforms such as Webtoons. This, in other words, is a clear example of how the development of the technical media of display affects the production and dissemination of qualified media types.

The integration of text and image in comics has indeed been a touchstone for first dismissing and then embracing comics as an object of study. The enmeshed use of images was initially used as a pretext to dismiss comics because they were not considered to be a serious medium and were thought to be unable to do anything beyond entertaining teenagers. This dismissive attitude may also have been a symptom of the difficulties comics posed to the disciplines of visual art and literature because both were unable to acknowledge and address the 'very adoption of pictorial information as narrative resource' (Bateman 2014, p. 91). It was with the publication of ground-breaking graphic narratives such as Art Spiegelman's *Maus* (published serially between 1980 and 1991; Spiegelman 1991) that the potential of the medium was brought into sight.

The graphic narrative is now a popular case study for both intermedial and multimodal studies due to its explicit foregrounding of the interaction between text and image. It is quite difficult to draw a clear line between multimodal and intermedial approaches to comics as they have a lot in common. The relation

between various semiotic modes in comics has been studied by scholars who situate themselves in the linguistic and multimodal directions (see Groensteen (2007), Cohn et al. (2012) and Bateman and Wildfeuer (2014)). Scholars working in literary studies, narratology and comics studies (see Chute (2008), Kukkonen (2013a, 2013b) and Stein (2015)) have also profited from using the terminology and analytical tools of intermediality and multimodality to discuss political, social and cognitive aspects of perceiving and interpreting comics.

Although words/texts and images are the basic defining elements of comics, they immediately lose their transparency when confronted on the comics page and the demarcating borders between the two, as agents of 'telling' (by words) and 'showing' (by images), start to blur. A comics page conventionally consists of basic media types of text and image framed within further non-linguistic elements such as colours, lines, panels, gutters, captions and speech balloons, which can as well be thought of as part of the 'image' element. Graphic narratives can also be wordless, as, for instance, is *The Arrival* by Shaun Tan (2007), which tells the story of a migration through a sequence of images unaccompanied by text, but wordless graphic narratives are not that common. Whether or not graphic narratives include text, all of their elements overlap at the material and sensorial levels as they are materialized altogether on the flat surface of a page, in print or digital, and are mainly perceived through the visual sense. When text and image are combined, they are closely integrated on the page in the material modality and we perceive them both visually in the sensorial modality. As will be explained in the rest of the chapter, it is mainly in the spatiotemporal and semiotic modalities that the differences between various interacting modes emerge. Before moving forward, however, it would be useful if you had a look at the technical terms in Box 8.1.

Box 8.1 Comics terms explained

Panel: an image on the page representing a single moment of action. One comics page might contain one or several panels that are separated from each other.
Gutter: the space between the panels.
Caption: words in a separate box that accompany the panels.
Speech balloons/bubbles: balloon-shaped images containing characters' dialogues that are connected to the character with a 'pointer'. **Thought balloons**, which usually look like clouds, contain characters' thoughts.
Sound-effect: words that simulate sounds and usually exist outside the captions or speech balloons.

Before delving deeper into these interactions, it is important to think about the process of 'meaning multiplication', which happens in the intermedial event that occurs when various semiotic modes join together. As John Bateman

(2014) defines it, meaning multiplication refers to new meanings that emerge from the interaction between different semiotic modes:

> Under the right conditions, the value of a combination of different modes of meaning can be worth more than the information (whatever that might be) that we get from the modes when used alone. In other words, text 'multiplied by' images is more than text simply occurring with or alongside images.
>
> (Bateman 2014, p. 6)

Meaning multiplication is of course not exclusive to word–image interactions in comics and is a principle that applies to all sorts of media combinations. In the specific case of comics, scholars have attempted to categorize different types of interaction between words and images, which create different added values in the process of meaning-making. Based on the level of integration or separation between the modes, or the dominance of one over the other, comics' styles are categorized as picture-specific (where pictures dominate and there are not many or even no words) or word-specific (where pictures are mostly additive or illustrative) or as having other integrated styles where the pictures and words have almost equal weight in the narrative and alternate to advance the story or mutually engage in doing so.

Often, as mentioned above, it is not easy to separate the basic media types of text and image and evaluate their weight as they merge together on the page and semiotically interact in a way that means we can say that words become images and images become words. In semiotic terms, that is, in terms of readers being conventionally used to words performing the symbolic function and images being mostly iconic and maybe to a lesser extent indexical on the comics page, these functions radically merge. Visual aspects of words become important through strategic usage of typography, and images, through their sequential repetition, convey meanings and narratives *just like* words. The semiotic interplay can go even further and summon sound through visual perception, such as by using bold characters in a speech balloon to imply that the words are being pronounced louder than the rest, or using sound effects and onomatopoeic images or words to represent the sound of a creaking door or a grumpy dog.

Let's take a look at a few examples. 'Space' (http://www.lunarbaboon.com/comics/space-1.html) is one of a series of single-paged webcomics picturing a character called 'Lunarbaboon', who is a half-man/half-monkey, and his daily life (Grady 2020). *Lunarbaboon* (http://www.lunarbaboon.com) comics are created by Christopher Grady (2012–) present) and they are published 2–3 times a week on a blog of the same name, and also on other webcomics platforms. The short graphic narratives, resembling flash fiction (extremely brief stories of no longer than a few paragraphs) have a minimal and readily graspable style and use a minimum yet clever mixture of semiotic modes. In Space, like other *Lunarbaboon* comics, you can see how the non-demarcated use of dialogue within the images and the style of the handwritten words capture the casual and ordinary elements of a father–daughter interaction over a day.

One interesting aspect of comics with which 'Space', (http://www.lunarbaboon.com/comics/space-1.html) plays is how temporality is constructed through spatial elements. Indeed, comics are a qualified media type that spatializes temporality, or, as Hillary Chute puts it, '[c]omics might be defined as a hybrid word-and-image form in which two narrative tracks, one verbal and one visual, register temporality spatially' (Chute 2008, p. 452). This representation of time and the implication of advancement in time is to a great extent performed by gutters. As Daniel Stein (2015) explains,

> This spatiotemporal construction – the representation of time through the techniques of panel design and sequencing – enlists readers to invest the gutters with meaning: to provide the links between panels by way of imagining what must have happened between one scene and the next.
>
> (Stein 2015, p. 424)

You can see that in the *Lunarbaboon* example, the gutter is foregrounded and played with to make the reader/viewer jump between temporal spots and become self-aware of the concept of time and its limitedness. The narrative is indeed told through panels 1, 2 and 4 and we are guided to jump over the third panel by the characters' hands. If we follow the pointing finger of one of the characters, this would mean that we would view the end of the narrative first (in the fourth panel) before looking at the third panel to glimpse several moments of the day in sub-panels. A diversified set of temporal concepts such as speed, a moment, past and present are represented like this through spatial elements and in the minimal and brief interaction between different semiotic modes.

Navigating through graphic narratives is not always an easy job and is called 'decoding' by scholars and artists. 'Reading' is obviously not enough because different types of signs with different dynamics of interpretation combine and merge and the mind has to be ready to make decisions and shift at any moment. To interact with comics, it is necessary to adopt a certain level of 'intermedial literacy' and be aware of the relations between modes to understand the intermedial relations.

Depending on the complexity of the relation between the words and images, different degrees of 'narrativization' efforts are needed to comprehend comics. Comics, with their complex engagement of semiotic modes, have been considered to work in some ways like films, but one of the main differences between the two media types is that in comics the narrative movement needs to be fuelled by the perceiver's mind. The perceiver has to move over the empty spaces of gutters while filling them with meanings and advancing the narrative. This process gets even more complicated when frames and borders are played with on the comics page.

In Figure 8.1, taken from *Fun Home, a Family Tragicomic* by Alison Bechdel (2007), you can see a sophisticated engagement with the medium via multiple layers of intermedial relations. As one of the most critically acclaimed and widely known graphic novels, *Fun Home* is distinguished by its complex

Figure 8.1 Fun Home (Bechdel 2007, p. 120).

autobiographical and contemplative wordy narrative and realistic style of drawing, which negotiates authenticity by representing and mixing different media. In the panel you can see a moment of 'remembering' which is happening through media representation, which is presented twice, one in the bottom panel and one in a more focused way in the top panel, as the narrator is looking at her father's photos and thinking about them. Although the words and images are quite neatly distinguished on the page, they create a complicated decoding moment, forcing the reader/viewer to pause, just like the narrator is doing, to get closer to or further from the object that is within view and to move between the images and their descriptions which are themselves, to some extent, transmediations of the images into textual descriptions.

Furthermore, the very act of mediation is foregrounded in the narrative and in the caption: the photos are viewed and read as incomplete 'translations' of what they have captured. This neatly shows the multiple layers of transmediation that are further intensified via the autobiographical agenda of the book: drawn photos on the page are suggested as representations of actual photos; those photos in turn are transmediations of actual moments and are parts of a whole that are being transmediated to ekphrastic texts and that is being commented upon on the comics page. What seems to be a media combination at first sight is actually made up of diversified sets of intermedial relations, including both media transformation and media representation.

Both of the examples discussed here have been recreated in new forms. *Lunarbaboon* was remediated as a book in 2017 and *Fun Home* was adapted to become a musical by Lisa Kron and Jeanine Tesori in 2013 and had widespread success. The medium of comics has proved to be popular for adaptations. It is now very common to see comics adapted to become feature films and TV series, especially in the case of pop-culture media products, such as *The Avengers, Batman* and *Wonder Woman*; it is as if comics invite adaptation – that the medium of film is asked for to fill the gaps in the gutters, so to speak.

Understanding filmic sound design. Or how raindrops can become a qualified medium

Sound film is frequently brought up as being an example of a medium that consists of intermedial combinations, because it so clearly depends on combining sound and image – two distinct basic media types. Yet it should be noted that when we watch a film, sound and image are *always* experienced together, and the potential for different meanings are interwoven to such a degree that the film should be considered more of an integrated medium than a combination (see Chapter 2). A great example of this is *sound effects*, which can be considered to be a qualified submedium in itself. Sound effects, as we will show in this section, contribute significantly to our understanding of a film, even when we don't think about sound consciously. As we will show, the images and the sound of, for example, a door slamming will often be *produced* independently of each other and then *combined* in post-production, but that doesn't

mean that we *experience* the door slamming in a film as a combination. We know from real life that a door slamming will cause a sound, and therefore the intuitive way for us to experience this event in a film would be to experience the sound and the image as an inseparable unit. There is, therefore, an analytical difference in whether a film is studied from the point of production or from the point of audience reception.

Keeping this integration and co-dependence of the medial forms in mind, it can still be useful from an analytical perspective to try and map out the different constituting media types of a film; that is, to look at how they are individually 'composed' and how they are combined to form an integrated audiovisual experience. In this section, we delve into the importance of sound effects for creating an audiovisual narrative. We first present a few key terms and functions of sound effects, which are then put to use in an analysis of a brief segment from the animated film *My Neighbor Totoro* (1988). Lastly, we present a short discussion on how to understand sound effects as a qualified medium in itself, which shows how we can think about the contextual and operational qualifying aspects of this particular medium.

What are sound effects?

Sound effects refer to all sounds in a film that cannot be classified as either speech or music, but the divide between sound effects and music is becoming more and more blurred, with sound and music often working closely together, or sound effects taking over the role of music altogether (cf. Kulezic-Wilson 2020). An example of the first instance can be seen in the intro sequence to *Corpse Bride* (2005), where 'mundane' sounds of a pen scratching and the noises of cutting a fish and sliding it off a counter are used both as a supplement and a contrast to the tune being played on the piano. Another example can also be found in the intro to *Atonement* (2007), where the tapping noises of someone using an old typewriter continue long after the writing is done and are integrated as a percussion element into the music that follows. Sound effects that are used 'in place' of music can be seen in *The Birds* (1963) in the scene where the schoolchildren and Melanie are fleeing from the school. A scene like this would conventionally be underscored with music to emphasize the terror and drama, but in *The Birds*, the sounds are electronically created sound effects that mimic the flapping and shrieking of birds, which combine with the children's screams to create the panicked atmosphere (cf. Wierzbicki 2008). These examples aside, the primary function of sound effects is to provide the sounds that objects, interactions and environments make in a film, such as slamming doors, blowing wind and footsteps on gravel.

Just like the electronic sounds in *The Birds*, most sound effects in contemporary mainstream film will be produced either as 'Foley sounds' or as 'library sounds', i.e. pre-recorded Foley or midi sounds stored in a digital library for later use. Foley sounds are effects that are performed by special Foley artists in a studio, where they use their own bodies and props to create sounds that

can replace or enhance selected sounds from the production track (audio recordings taken simultaneously with the visual recordings), such as the noises of footsteps or horses galloping (the classic example is to mimic this sound using coconut shells). Very often, the sound that is recorded while filming is not suitable for the film. There might be too much noise on the recording, or the sound that has been recorded might not live up to filmic conventions dictating that some sounds should be exaggerated or modified compared to real-life sound (think, for example, of gunshots and explosions). Also, the film might need sounds for things such as dinosaurs, which do not exist in the real world. Finally, it might be necessary to be able to replace the dialogue (e.g. for films with synchronized dialogue in different languages), and for this purpose, it is more effective or practical to use Foley, which can be recorded and manipulated independently.

This use of Foley and library sounds means that the production method used for sounds in live-action film – at least expensive mainstream film – and animation doesn't necessarily differ significantly, except in relation to films that require extensive use of production sound because of specific genre conventions or budget restrictions, as is the case, for example, with the Dogme 95 films, which do not allow Foley to be used.

Sound effects as mediators of auditory information

In sound studies, the sound of our environment is often referred to as a 'soundscape' that is made up of three different types of sound: 'keynote sounds', 'signal sounds' and 'sound marks' (Schafer 1994 [1977], pp. 9–10). The keynote sound is the 'background sound' and can, for example, consist of waves on the beach, birds singing or traffic noise. It is the sound that you don't consciously pay attention to. The signal sounds, on the other hand, are the sounds you do pay attention to, like a sudden cry of a seagull breaking the relative 'silence' of the waves. The last type of sound is the sound mark, which is a sound that is uniquely tied to a certain place. Church bells will most often be signal sounds, but in cities where the bells play unique melodies, this can be a sound mark.

Sound effects are basically used to create *filmic* soundscapes, using the iconic mode to refer to an extra-filmic reality (as also discussed in the section on radio drama later in this chapter). Soundscapes could be described as the unorganized basic medium that sound effects, a highly organized qualified medium, builds on and imitates. We, therefore, suggest using the terms keynote sound, signal sound and sound mark as one way of classifying and approaching sound effects in film and focusing the analysis on how this filmic soundscape is constructed. The next step is to analyze *what* the individual sounds, in their functions of keynote, signal sound or sound mark, are used to portray about the atmosphere, environment, objects or characters in the scene. We go into more detail next about how these different types of sound add meaning to a film.

The keynote sounds are fundamental for setting a scene in a specific location, while at the same time also contributing to the level of tension in a scene: the

sound of an ocean can, for example, tell us that we are dealing with an event that is happening close to the sea, even if the sea is not visible in the shot, at the same time as it creates a peaceful and idyllic atmosphere. A typical use of a keynote sound in film is to use the sound of rain or thunder when lovers in a romantic film are in distress. The sound relies on both the iconic and the symbolic mode for creating meaning – iconic because it resembles sound from our everyday lives, and symbolic because rain and thunder are connected with emotional suffering by convention. An example of this is the scene from the 2005 film *Pride and Prejudice* (2005) when Mr Darcy proposes to Lizzie only to be rejected after a bitter fight. This scene is aptly anticipated by dramatic violins *and* by the sound of thunder. During the two characters' conversation, the music stops to leave room for the dialogue, but the keynote sound of rain and thunder continues and functions to provide a background atmosphere.

Both keynote sounds and signal sounds also provide concrete information about objects and characters in a film. When we listen to sound in our everyday life, we, more or less subconsciously, decode sound to obtain information about the objects and interactions that have created it. Sound is always 'shaped' by the physical properties of its source (Gaver 1993), which makes it possible for us to pick up information about the materiality and placement of a sound source. We use this ability to decode auditory information actively when we knock on a wall to determine its material and thickness before deciding whether to drill into it. Sound thus adds information to the film about the material and spatial qualities of objects and environments, information which is not always provided by the images.

Finally, sound marks will not always be relevant in a film, but you might find examples of films in which a specific sound is associated with a specific place and which thus functions as a symbol of that place.

Even though sound effects are very often exaggerated Foley or library effects, these unrealistic sounds still share characteristics with the sounds they are supposed to be iconic signs of, and if the exaggeration is done in accordance with the specific conventions of a film genre, we tend to accept them as realistic without thinking about it (Langkjær 2010). In this way, sound effects not only gain their meaning through the iconic mode but also through the indexical mode by having a direct relation with the filmed materials (production sound), or simply by having an *imagined* relation to this material (Foley and library effects).

Sound can also be used in an explicitly unrealistic way, however, commonly for comic relief or as a strategy for pointing to the film as a constructed representation. This is a popular strategy in some genres of animation, such as when objects in *Looney Toons* are matched with incoherent sound.

To sum up, one way to approach the analysis of sound effects in film is to ask yourself how the soundscape is composed of keynote sounds, signal sounds and sound marks. What is the function of the keynote sounds? Which signal sounds are designed to stand out, and to what effect? Second, you might ask yourself what the style of the sound effects is. Do they aim to represent or exaggerate what is visually represented or to create an abstract effect, as in

Loony Toons? Other questions to ask are: To what degree do the sound effects represent objects or features, such as thunder, that are not visible in the images? What information about space, place and materiality is provided by the sound How does this auditive information work with the information provided by the images, and how are the sound effects synched with elements in the images? The sound effects might even emphasize something in the images which you otherwise would not have paid attention to. Finally, you could ask to what degree the sound effects are integrated with, or are being substituted for, the use of background music.

This list is not meant to be the ultimate model for analysis, but an inspiration for how to begin to think critically about the function of sound in relation to the other media in film, which we will exemplify briefly next.

Sound and the subjective listening position in My Neighbor Totoro

Many of the aspects of sound that we have detailed above can be exemplified with reference to a short clip from the animated feature film *My Neighbor Totoro* (1988). The film is about two girls, Satsuki and Mei, who, afraid and scared while their mum is in hospital, find comfort when they meet a new friend, the fantastical creature Totoro. The scene we wish to discuss occurs about 50 minutes into the film and lasts for a little less than a minute (00:49:42–00:50:32).

Totoro is standing by the bus stop with the two girls; it is raining, and Satsuki has just handed Totoro an umbrella. When Totoro puts the umbrella over his head, he has the experience of hearing rain on an umbrella for the very first time, and this whole scene is a little poetic interlude which shows how different rain can sound depending on the material it interacts with, the listener's position, and the listener's attitude to the rain. In this short clip, there are at least five different sounds of rain: 1) a steady rain sound used as a keynote, characterized as light rain hitting the ground and landing in puddles, 2) the sound of soft splashes of raindrops hitting the head of a frog, 3) the sounds of hard, hollow-sounding drops landing on Totoro's umbrella, 4) the sound of heavy rain hitting the ground far away, and lastly 5) the sound of a cascade of heavy raindrops on the umbrella heard from a short distance away rather than from directly underneath the umbrella.

Although the scene is created by 'cutting' between images of the ground, the umbrella and the frog, raindrops are hard to truly represent only in images. The sound here is (co-)creating our impressions of the size and amounts of raindrops, the placement of the listener in relation to the impact of the raindrops, and what material the ground and objects that the rain is falling on are made of.

Besides giving us all this information about materials and listening position, this scene also shows us something about listening attitude. The first type of rain sound is characterized as a keynote sound, it is continuous without really calling attention to itself, and it keeps playing underneath all the other sounds.

But, as soon as the filmic focus shifts to the different types of raindrops in the next few cuts (the integration and synchronization of sound effects and visual editing is significant here), the sound of raindrops has become the primary focus. The individual drops are signal sounds now that we, along with Totoro, listen to actively. In this way, we are sharing Totoro's auditory attention and wonder concerning the diversity and musical experience hidden in a mundane thing such as falling rain.

Is sound design really a qualified medium?

It should be apparent from the account we have given so far that sound effects are used differently in different genres, and that sound effects are designed both to communicate certain information and to have an aesthetic effect, both of which are subject to conventions. Because sound effects are not just the pure resource of available sound, but are, on the contrary, specifically shaped and moulded to suit specific needs and conventions, we argue that sound effects, or perhaps more accurately *sound design*, is, in fact, a qualified medium with its own creative history.

Foley sounds have existed since the very early sound films, and were used for the first time in the musical *Showboat* (1929) (Theme Ament 2014). As the technology became more advanced, so did sound effects, and Foley and sound editing became a substantial part of film in the late 1960s, and a dedicated Oscar for best sound editing has existed since 1963 (sound editing, here, refers to the selection and placement of sound). From 1930, however, the Academy has been giving out an award for *sound mixing*, that is, for adjusting the volume and other parameters of the sounds on the soundtrack to make it all work together as a whole.

It is clear that sound design, which we use as an overarching term covering the composition of sound effects as it exists in a film (after both editing and mixing), is partly constrained by the technological development but also by the *institutions* that exist around sound design. Besides the Oscars presiding over what a 'good' sound design is in the Hollywood movie, sound design is, and always has been, something that has been taught either through apprenticeship or through specialized programmes at universities or film schools – to which the growing body of specialized literature also speaks. The industry has, furthermore, seen a number of creative individuals who have helped shape the practices and conventions. A few names to highlight here could be Julien Naudin (collaborates with Lars von Trier), Aldo Ciorba (collaborates with Sergio Leone), Ben Burtt (*Wall-E* and most of the *Star Wars* films) and, not least, Walther Murch (*Apocalypse Now*).

The existence of an institutionalized tradition of development as well as role models who serve to regulate sound design to follow communicative and aesthetic conventions means that the requirements of the contextual and the operational qualifying aspects are met – making sound design a qualified medium.

Media combination in radio drama. The integration of sound and music in *The Unforgiven*

We live in a world consisting of narratives of different types: novels, films, films based on novels, TV series based on films, documentary dramas, game plays, etc. Apart from the obvious intermedial aspects of these examples, as media combinations and media transformations, what most of these media products have in common is their dependence on the visual mode in the sensorial modality. One consequence of this presence has led to narrative studies that have largely focused on the 'verbal/textual in combination with visual or audiovisual media', which in turn has created blind spots within the general study of narratology (Mildorf and Kinzel 2016, pp. 1–2). This is the case for radio drama – a sound-based qualified medium. As Hugh Chignell, for instance, asks: how is it possible in a visual culture for 'invisible' drama to exist? (Chignell 2009, p. 26).

The answer might just be that it does exist but that due to its 'invisibility', not enough attention has been paid to it. Radio drama has barely been discussed within intermedial studies until quite recently, and when it has been studied, the view has been that aspects such as the function of sound and music within it are inferior to the narrative structure. In other words, radio drama has been regarded as a literary genre rather than what we are emphasizing that it is here – a sound-based qualified media type (Lutostanski (2016, p. 117), and also see Mader (2007, pp. 179–83)) or a submedium of audio narrative (see Huwiler (2016, p. 99) and Mildorf and Kinzel (2016, pp. 8–9)).

The purpose of this section is thus to 'upgrade' radio drama as a sound-based qualified media type that is equivalent to other narrative media types such as films and TV series by focusing on sound and music but without compromising other key aspects that are typical of narratology, such as narrative events, the relationship between diegesis and non-diegesis, dialogues and monologues, and focalization (Bernaerts 2016, p. 133). This section will discuss a few excerpts from the five-episode crime-fiction radio drama *The Unforgiven* (2018), a so-called prequel to the Emmy Award-winning crime-fiction TV series *Waking the Dead* (2000–11), which investigates cold cases in and around London.[1] A prequel, it should be noted, is, according to the *Oxford English Dictionary*, a 'book, film, etc., narrating events which precede those of an existing work'.[2] This means that the genre (crime fiction) has already been indicated to the listener and that the main characters from the TV series are recognizable within the radio drama. This relation thus frames the listener with a number of given cultural and semiotic codes that are provided beforehand. But what characterizes a radio drama? What are its basic constituent elements?

Radio drama: A hybrid and multimodal narrative art form

Radio drama is a 'hybrid and multimodal form' (Bernaerts 2017, p. 206) in which the story unfolds in time and space through perceived intricate

interactions between basic and qualified media such as speech, music, voice, sound and silence. For instance, sound appears in speech in dialogues and monologues as well as in the speech of different types of narrators, not least in the tone of the voice that is central to understanding the characters' different emotional states (Huwiler 2016, p. 103). Dialogue is meant to mean 'talk-in-interaction', just as it is used in everyday life (Mildorf and Thomas 2016, p. 3). The dialogues and sounds thus function to create an effect of realism. Furthermore, on the one hand, sound creates spaces by referring to real-life objects, such as slamming doors or car engines (sound marks). On the other hand, it 'creates spaces of action', such as when an action takes place in a car and that action is marked by the sound of a car engine (soundscape). Sounds may also create 'routes' between different places – here the mixing of sounds and microphone placement function as a framing focalizer for the listener (Lutostanski 2016, pp. 120–1). Sounds also appear in non-diegetic sound effects.

Music functions to underscore specific narrative events, emphasizing emotional states and creating a specific atmosphere, and forms part of the diegetic world. In this sense, radio dramas resemble films. Moreover, words and sounds in combination may create a sense of visual imagery by way of suggestion (Allen 2008, p. 481). Finally, radio drama is also a special type of (technological) mediatized performance, where the actions sometimes take place at the same time as they appear to the listener, either by way of 'showing' or by way of 'telling' what is happening. The way the story is driven, either in the form of dialogues or monologues, thus tells the listener in what mode the narrative events appear. In this sense, radio drama is a typical example of 'liveness' (Huwiler 2016, p. 106).

In the next sections, we highlight four types of functions that sound and music have for the narrative structure in the specific radio drama *The Unforgiven*: 1) sound as a visual marker, 2) words and sound in combination, 3) the use of diegetic and non-diegetic music and 4) verbal descriptions of sound in combination with sound.

Sound as a visual marker

Creating visual imagery by way of words occurs frequently in novels. But how can the same effect be created through sounds? While words are symbolic in character, sounds are most often iconic. As previously mentioned, the function of sound is to create visual imagery through suggestion; this can be done by using original sounds or by mixing sounds and microphone positioning. The issue in these cases is how to become aware of the visual imagery (sound mark), but also what kind of circumstance (soundscape) makes this significant. According to Werner Wolf (2016), the creation of visual imagery by way of words can 'only be perceived in relation to a real or expected presence' (p. 6). In the case of radio drama, there is no visual representation. However, we can explore how auditory basic and qualified media types, sounds, dialogue and music connect to visual imagery. Next we have an excerpt from the title

sequence of *The Unforgiven*, which illustrates this. It begins with the listener hearing seagulls in the sky and a police radio in the distance, and then there is a 'speech' by the Head Officer, Det. Supt. Peter Boyd:

> We all know the stats. If a murder doesn't get cracked in the first week, it's probably not going to happen. Unsolved crime. Cold cases. The Met's got more than its share. Which is where we come in.

The absence of visual representation makes the listener particularly aware of the surrounding sound effects. The seagulls and the police radio set the scene in a specific spatial environment and place the listener within a fictional world of crimes: the sounds of the seagulls and the police radio make the listener understand that the story is happening somewhere near water and near a crime scene. The sounds therefore become conventionalized as symbolic signs through their iconic nature, that is, the fact that they represent parts of reality. But they are also examples of indexical signs because of the absence of images – the sounds are pointing to something that is not visually present; that is, we cannot see it with our own eyes. The specific circumstance, the narrative event, makes this significant. This shows at least one difference between words and sounds in terms of the creation of visual imagery: the former do so through symbolic signs and the latter through iconic characters, thereby creating a specific soundscape.

Words and sound in combination and non-diegetic music

The title sequence goes back and forth between two different spatiotemporal settings: between a pub and the crime scene, between dialogues and Boyd's speech. The pub appears as a place for reflection and a 'talk-in-interaction', which is then transformed into the story, where the temporal and spatial distance between the event's actual appearances becomes increasingly present. Musically, the title sequence starts with a single stroke on a cello before we hear a police radio, seagulls and a humming sound. The scene then cuts to the pub, where we hear glasses and voices that are underscored with almost inaudible strings, then 'stitches' back to the first scene, when Boyd is continuing his speech. The speech testifies to a serious crime, which is heard through the tone of his voice. It captures the listener's attention, and the atmosphere is enhanced by dynamically increasing, non-diegetic music.

The story is told through dialogues and monologues from the perspectives of the different characters as if the events had already taken place. This first becomes clear at the pub when Boyd makes a comment in the conversation, saying, 'Hang on, this was my story', then continues when the title sequence has ended; this is marked with a three-note musical motif:

> 15 years before. Day one. Detective Sergeant Peter Boyd's story. February 10th, 1984. My house. My home. Raw winter's day. Sleet dropping out of a snow-filled sky.

So the story begins, and we are displaced in time, to 1984 and to 'a world without DNA, CCTV, mobiles, the internet and databases'.[3] Then there is a call from the commissioner saying that he wants to see Boyd about an old case, about a rapist-killer, a cold case that has now turned into a case of police corruption: Boyd is being accused of framing a killer who is already in prison for raping and killing five girls; the fifth girl's body never was found, but the killer claims he didn't kill her.

As suggested, the story is partly told through monologues. This telling mode often creates a sense of what is happening for the listener, but at a distance. The advantage of this distance is that it can be used to describe the diegetic world in detail, and the showing mode can't always do this. However, sometimes the telling mode merges with sound effects as if the distance to the listener has decreased and the story is happening in the present. An example of this is when Boyd takes the car to the police headquarters to meet the commissioner:

> It was snowing hard. Ice on the road. I enjoyed the risk. I forgot about Jen. My heart raced. There were days despite it all, you're like an athlete on the blocks, the race mentally run. Ahead of the pack coming off the bend, already a winner.

The visual scenery that emerges through the words uttered by Boyd is reinforced by sounds that iconically represent parts of reality. For instance, when Boyd says that it is snowing hard and that his heart raced, we not only hear a car driving off and honking horns, we also hear strong, pulsating music. This creates a spatiotemporal sense – the visual space of snow falling and the time of moving forwards – almost simultaneously. We also hear the sound of an icy wind in the air when he describes how the sky is full of snow. The combination of words and sound frames the whole story's mood here – a cold, barren everyday life. It creates a sense of visual imagery.

Diegetic and non-diegetic music

In episode one, the unit is driving off to an address in London, and this is marked with a comment by one of the unit's members, Spencer, 'OK. Let's go'. It is underscored with a short drumbeat just before we hear a car driving off at high speed. We also hear a voice from either a police radio or the car radio before the music is turned up – it is Maurice Ravel's ballet *Boléro* (1920) with its steady 3/4 rhythm. The other unit members, Grace, Mel and Frankie, are commenting on Spencer's musical choice: 'I didn't know you liked classical music?' 'I don't, it's my fiancée'. Then the music foregrounds as a kind of underscore to this narrative event, blending with sounds representing parts of reality. Since there are no clues through moving images, the microphone placement becomes even more important: it has become a kind of focalizer. In this case, the verbal comments on the music in the car also make the listener understand that the music belongs to the diegetic world. Here, media

transformations appear on different levels: partly via basic and qualified media types (sound, words, music), partly via a technical medium of display (car radio). The combination of these media types and the different modalities also creates a sense of spatiotemporality by way of the semiotic modality: the narrative event is moving forward towards the next event.

Verbal description of sound in combination with sound

While the intrigues replace each other with new evidence and new problems, such as a videotape that is delivered showing one of the cold case unit's daughters being abducted. The abducted daughter is holding a newspaper in her hand, and this sends a message to the cold case unit. The information that these (audio)visual media products provide in the diegesis is conveyed primarily by a verbal description of what the characters see on the videotape. As listeners, we don't hear all the diegetic sound elements of the video; instead, they are verbalized through a dialogue, but there are sounds that can be heard. In the following quote, the transcription focuses on the dialogue.

FRANKIE: Kate's [the daughter] video has a background noise, a steady hum.
SPENCE: OK, that's traffic and it has to be moving above 30 miles an hour to create a hum, so we are looking for somewhere near a major road.
MEL: OK, so where the flats run apart there's no big roads.
SPENCE: OK, they have vanished.
FRANKIE: The audio report also picked up an intermittent low-level rumble
SPENCE: Got to be the Tube, surely.
GRACE: Yes, yes.
SPENCE: Where's that on the map?
MEL: OK, I overlay the tube lines. There, Hammersmith and District.
GRACE: And, it doesn't go any near those flats, good.
FRANKIE: There's also a screeching noise near the end of the video.
SPENCE: Let me hear it again.
FRANKIE: And it's a train braking, surely.
GRACE: Does it come to a stop, so held at signal, end of a journey, near a station?
SPENCE: Fast road, near train tracks, over tube, that's this area here.
MEL: Well, the road could be the Westway flyover?
FRANKIE: Yes, that's close enough to the Paddington line. Check the signal points Mel!
MEL: OK, ah, here and here, so, it narrows to round this section here.

When there is a description of the train braking, the listener also hears the sound of something that iconically represents a train braking. What is prominent in this example is that the sounds that the listener hears are followed by verbal descriptions of visual perception: as listeners, we 'see' a train in front of our inner eyes through the combination of words and sounds. But not only

that, the verbal descriptions of the sounds also lead the listener to possible places in London – the narrative event is thus an example of how the spatiotemporal modality is made 'visible' to the listener through the semiotic modality. The sounds become narrative agents within the diegetic world.

Although radio drama only consists of sound, the example discussed in this section shows how complex a qualified medium it is and can be. Based on auditory perception in the sensorial modality, the basic media and qualified media types of dialogue, sound effects and music interact in the auditory narrative. Dialogue and sound effects convey visual imagery and mediate between different narratives levels; music is used to create atmosphere. The different ways in which the different media types are combined in radio drama make it a particularly suitable subject to study from intermedial and narratological perspectives.

Songs and singing

Songs are probably the most widespread form of music and definitely one of the most widely consumed art forms. Songs mix words and music and can be mediated in sound or in writing. Singing is the performance of musical sound by the human voice; it is thus an *act* rather than an *artefact*. When analysing a song, we can choose between different medial representations, but the choice has implications for our results. If our analysis is based on a written song, we cannot consider aspects such as tone, timbre, sound effects or visual effects or qualities such as the nasal tone of Rihanna's voice or the guttural tone of Shakira's. If we are making an auditory analysis, we might have trouble hearing the lyrics. As songs are auditory in the sensorial modality, there will be a greater focus on the temporal mode because once we press play, the song starts unfolding in a fixed sequentiality. Even though we can of course rewind, the analysis is being done under wholly different conditions from analysing writing on a piece of paper. In the latter situation, we decide the reading pace and can go back and forth as we please.

The most crucial element of an intermedial analysis of a song is that of interpreting the relationship between the words and the music. Words and music share the same material, sensory and spatiotemporal qualities (see Chapter 4); they only really differ with respect to the semiotic modality in which words tend to make use of the symbolic mode and use the iconic mode less often than music. Some of the most palpably shared features of music and words are rhythm and tonal pitch, which is a good starting point for an intermedial analysis: how do the rhythm and melody of words and music interact? Likewise, the analyst might pursue the question of how music and words, with their different semiotic resources, work together to construe meaning in a song.

The analyst should avoid the common tendency to reduce the meaning of a song to its words. In everyday discourse, the question of what a song is 'about' generally refers to what the lyrics are about. When commenting on the music of a song, we often say that music emphasizes or highlights the text. In the words of musicologist Nicholas Cook (1998), this terminology is dangerous

because we thereby 'imply that the music is supplementary to the meaning that is *already* in the words' (p. 54, emphasis in original). An intermedial song analysis needs to steer clear of this trap. One of its most important tasks is to show how the meaning of a song emerges from an *interaction* between words and music and that this interaction is emergent, that is, irreducible to the sum of its parts (Agawu 1992).

Finally, our analytical choices will vary greatly depending on the particular song genre in question. Classical art songs often lend themselves to a visually based analysis because much of their essence or authority as artworks is perceived to reside in their written appearance. Popular song genres such as rock and pop, on the other hand, are much more tied to the medium of the sound recording and thus, in relation to these genres, auditory aspects will often play a larger role in the analysis. These distinctions will be kept in mind in the next sections when we demonstrate how an intermedial analysis of a song might unfold.[4]

Interaction of words and music in a classical art song

German composer Franz Schubert's setting of a German translation by Adam Storck of a Walter Scott poem ('The Lady of the Lake', 1810) known as 'Ave Maria' presents an example of what Cook terms 'conformance' between media; that is, words and music join forces when they are communicating the same message or narrative (Cook 1998, p. 100). Properly titled '*Ellens dritter Gesang*' ('Ellen's Third Song'), this song is part of a song cycle, that is, a musical setting of a number of poems that have been set to music and that tell a consecutive story, a popular format in the nineteenth century (see Mandyczewski 1895, p. 90). If we take a look at the interaction between the words and music in the first stanza of the song, we can see how the melody is constructed to embody the meaning of the words (My own translation of Storck's text comes right below).

Ave Maria! Jungfrau mild,
Erhöre einer Jungfrau Flehen,
Aus diesem Felsen starr und wild
Soll mein Gebet zu dir hin wehen.
Wir schlafen sicher bis zum Morgen,
Ob Menschen noch so grausam sind.
O Jungfrau, sieh der Jungfrau Sorgen,
O Mutter, hör ein bittend Kind!
Ave Maria!

Ave Maria! Maiden mild,
Listen to a maiden's prayer,
From these rocks rigid and wild
My prayer shall to thee blow.
Safe may we sleep until the morning,

Though people are so cruel
Maiden, hear a maiden's prayer,
Mother, hear a suppliant child!
Ave Maria!

The poem is a prayer that is uttered by Ellen Douglas, 'the Lady of the Lake'. The context is that Ellen and her father have gone to hide in a mountain cave to escape a battle, and she sings this song to urge Mary to help and comfort them. The first line of the stanza containing the anaphoric appeal to the virgin is set to a calm, simple melodic line with straightforward chords. This depicts the confident and hopeful tone of Ellen's prayer. Further, Schubert's composition stretches the words 'Ave Maria' to last more than a whole bar. The two stressed syllables (**A**ve Mar**i**a) are thus significantly prolonged, which makes the listener dwell more on these words than probably would have been the case during a reading of the bare, written poem. The second line, in which Ellen's desperation increases, opens with a gloomy augmented chord that unsettles the hitherto innocuous harmony. The wild and rigid rocks of the third line are portrayed by a tortuously winding melody. In the fourth line, in bar eight, on the word 'blow', (*wehen*), the melody elopes from the words in a long melisma, depicting how Ellen's prayer is blown to the Virgin Mary.

A melisma is a stretching of a single syllable over several musical notes. Whenever a melismatic setting occurs, it has the effect of emancipating music from words in that the melodic movement is conditioned by musical form, not verbal progression. Conversely, syllabic form, in which each syllable corresponds to one musical note, presents a declamatory, verbatim and word-centred compositional style. In this specific case, Schubert lets music escape from the words just at the same time as Ellen's prayer leaves the cave. In the fifth line, Ellen's prayer that they may sleep safely is rendered in a soft, swaying melody that is reminiscent of a lullaby. In the sixth line, the major key shifts to a darker minor one at the same time as Ellen talks about the cruelty of the world. The seventh and eighth lines articulate Ellen's pleading with yearning appoggiaturas (a traditional way of musically expressing longing) on the words 'Sorgen' and 'Mutter'. Finally, the last line brings us back to the comforting state of the opening on the words 'Ave Maria'.

Our analysis demonstrates the close alignment between words and music. At the level of the semiotic modality, Schubert's music provides iconic representations of the narrative that is communicated via symbolic references in words. For instance, the waving melisma, with its rapid tones on the word *wehen* iconically represents a whirling wind. This compositional style, with music iconically miming words, is common in many classical genres, not least in the motets and madrigals of the Renaissance and Baroque styles. In the late Romantic period, we increasingly see examples of music detaching from words and becoming more independent. A poignant example is the last song of Heinrich Heine's and Robert Schumann's *Dichterliebe* (1840), an ironic song of a depraved lover that ends with a musical coda that is in a totally different, mild and optimistic tone from the rest of the song, suggesting a wholly different ending to the story than that of the words.

Another crucial issue in intermedial analyses of songs is discussing positions of enunciation; that is, the question of who is communicating to whom. Pop music scholarship has identified three 'personae' at play in a song (see Frith (1996) and Auslander (2004)); for example, whenever Lady Gaga sings a song, she is simultaneously a real person with a civic name, a culturally and commercially constructed persona who is a star with a stage name, and a character in the narrative of the song. In the case of 'Ave Maria', Ellen is a character in the story of the dramatic poem of which the song is part. She thus stands as the communicator of the uttered words, with the Virgin Mary as the recipient. But the person who wrote the words is obviously in some sense also their utterer – and in this case, we are dealing with a translation, so both the translator (Adam Storck) and the original poet (Walter Scott) must be seen as uttering a message to us in the poem. Then there is the composer – by musically interpreting the words, he is adopting them and making them 'his', so much so that in classical music, we simply refer to 'Schubert's songs'. Finally, there is the performer, who also has some claim to be viewed as the communicator – and in this case we, members of the audience, are the recipients. When a singer sings 'Ave Maria', she is in a sense simultaneously both herself singing to an audience *and* Ellen praying to Mary. In classical music, the same song can be performed by several singers without making one of them 'the original' and the others 'covers', like in popular music. This means that every time a singer sings Schubert's song, she is acquiring it and making its words her own.

Beyoncé's 2008 reworking of 'Ave Maria' retains the 'refrain' and part of the musical accompaniment. This version makes the enunciation structure of the song even more intricate, as Beyoncé's lyrics address both a 'you' and an 'I', but they are very different ones from those in the original poem. In Ellen's prayer to Mary, the personal pronouns are referring to characters in the story (and to the mythical persona of the Virgin Mary); in Beyoncé's version, we can probably accurately speculate that she is addressing a very real person, namely her husband, Jay-Z. The question of who is communicating something to us in a song is complex and involves a vast array of personae.

Contest between sound and sense in a pop song

While 'Ave Maria' is an example of medial conformance between words and music, let's also consider an example of the opposite. The music in a song does not always merely support the message of the lyrics; in some cases, the music and lyrics can seem to contradict each other. Consider these words:

> Oh, we've been together
> But separate is always better when there's feelings involved.
> If what they say is 'nothing is forever',
> Then what makes love the exception?
> So why, oh why are we so in denial
> When we know we are not happy here?

Read as bare text, we are likely to interpret these words as part of a bitter song by someone or about someone who is heartbroken. But when we consider the song as a whole, including its music, the picture changes radically. This song, 'Hey Ya' by the hip-hop group Outkast, has cheerfully upbeat music and is set in a major key. The lead singer, André 3000, sings energetically at a fast tempo and in a bright, slightly nasal tone. The music of the song communicates happy feelings and probably makes listeners want to dance. Music in this instance communicates virtually the opposite of the lyrics, and if we based an analysis on the written words alone, we would come to a conclusion about the meaning of the song that lies far away from the lived experience of listening to it.

Since music and words convey conflicting messages here, we are witnessing what Cook terms a medial 'contest' (Cook 1998, p. 103). In Cook's view, such instances are in fact the most interesting from an intermedial perspective, because the different media involved in the utterance do not just replicate each other but create new meaning through their collision. In the case of 'Hey Ya', we could thus interpret the meaning of the song as emerging from the clash of the melancholy of the lyrics with the optimism of the music: the music provides hope despite the hopelessness of the lyrics. Or perhaps the cheerful music could be heard as an ironic commentary on the disillusioned words. However, we need to consider that the lyrics to pop songs like 'Hey Ya' usually have no independent existence outside the sounding music. The words can thus easily be dominated, outshouted, as it were, by sound. If that is the case, we can say that the material modality exceeds the semiotic in importance. Indeed, it is not even certain that the majority of listeners will actively listen to the lyrics. As such, if there is a contest between the qualified media of music and lyrics here, music is likely to have the upper hand.

Moreover, the lyrics to 'Hey Ya' are full of repetitions. The penultimate line in the above excerpt would be more correctly transcribed as 'So why oh why oh, why oh why oh why oh, are we so in denial […]?' The repetition of words or phrases leads us to focus on their auditory form rather than on their meaning. The sound of the diphthong 'why oh' only becomes more evocative as it is repeated again and again rhythmically. All the while, the actual question posed here, 'why are we so in denial?', is becoming more and more obscured. Even the subsequent words become sucked into this black hole of non-semantic sound, since 'denial' rhymes with 'why oh'. Rhyme is one of the oldest tricks in the book that can be used to make words more music-like, because rhyming words, like music, exploit the aesthetic quality of sound. Rhyme and repetition are thus devices that draw words towards the medial domain of music. We are used to thinking of the sound of the voice as providing building blocks for uttering meaningful words. But in the case of 'Hey Ya', we encounter the opposite logic: words are employed as sounding building blocks, in line with instrumental sounds and sound effects, and their symbolic meaning subsides.

This effect becomes increasingly pronounced as 'Hey Ya' progresses; towards the end of the song, word repetitions and non-semantic interjections seem to

take over, such as when the word 'alright' is almost infinitely repeated or when the phrase 'shake it', syncopated and with great emphasis on the aspirated consonants, becomes a percussive element. It is as though the semantic meaning of the lyrics is gradually dissolving itself as the aesthetic quality of sound takes over. We have a similar effect in other pop songs (Borčak 2017). In Rihanna's 2007 song 'Umbrella', the word 'umbrella' is gradually deconstructed in the course of the refrain, ending in just one repeated syllable that serves only as a means for rhythmically displaying Rihanna's bright, piercing voice: 'umbrella–ella–ella–eh–eh–eh'. These typical traits of pop lyrics are also indicative of the fact that, unlike classical art song lyrics, they are not intended to gain aesthetic attention outside the environment of the sounding music. We could make a full-on literary analysis of Walter Scott's 'Lady of the Lake' because the poem existed prior to being set to song. Although Schubert's music greatly popularized the text and probably saved it from oblivion, the poem has coherent meaning in and of itself. The same cannot be said about pop song lyrics.

Words and music – and beyond

One of the ways in which music creates meaning, apart from including iconic representations of movements or feelings, is by having intertextual references. In Stromae's song 'Carmen' (2015), the lyrics are about how social media are consuming our attention and impounding our social life. Since Stromae sings and raps exclusively in French, it is likely that a large part of his international audience do not understand the meaning of the lyrics, but they get considerable help in this respect from two other involved media: music and video.

The music draws heavily on the famous Habanera aria from Bizet's opera *Carmen* that is sung by the femme fatale Carmen, who leads men astray only to deceive them. The aria titled 'Love is a Rebellious Bird' is about how love cannot be tamed. The intertextual use of the suggestive musical theme is therefore associated with a narrative about seduction and deception; that is, if the listener is well-enough acquainted with a standard Western music repertoire to appreciate this reference.

Both the lyrics and the music in this song thus require specific knowledge (mastery of the French language and/or the Western classical music canon) to be fully understood. Far more palpable and something that has a broader reach is the message communicated visually in the music video, in which a little blue bird, recognizable to most people as the Twitter symbol, becomes increasingly larger as it consumes the social life of the main character until it eats him and others up and then poops them out at the end.

'Carmen' can be considered an example of Cook's third and last type of intermedial interaction: complementation; that is, the different media involved are responsible for different parts of the meaning-making process (Cook 1998, pp. 103). The lyrics warn that social media can be time-consuming. The music adds a dimension that shows that social media are like the rebellious and

deceptive bird in Habanera's aria. Lastly, the video visualizes how the negative impact of social media in our life becomes larger and larger the more we engage with them. The importance of the video in this song exemplifies the decisive role that the medium of the music video plays in a great many song genres. The next section will explore the music video in depth.

The intermedial music video

Throughout history there has been a tendency to add something to music, or to perform music in combinations with art, bodies, language and other forms of media. Consequently, when human beings have experienced music of different kinds, there has always been something more than the music to experience. When music is performed, we, the listening and watching audience, see the musicians and their instruments, we might hear other (non-musical) sounds, we see all kinds of visual material, architectural space and we feel the music physically moving through the air and into and onto our bodies, ears, etc. The visual senses, in particular, have been 'the extra' sensorial modality being connected to music as its medial 'package'. This goes for religious as well as non-religious situations, for political, cultural and aesthetic contexts in which music has played an important role through the centuries.

Richard Wagner (1813–83) confirmed his ideas about opera and the *Gesamtkunstwerk* (meaning the 'total work of art'), which should include all art forms and invite the audience to take part in the production of meaning and in the aesthetic experience (Wilson Smith 2007). Later on, different media technologies made it possible to record and store sounds and visual media, and also combine them (sound film appeared as an invention in the late 1920s). Records (and record sleeves), film music, commercials, television and games form the modern mediascape, and today we take it for granted, perhaps more than in relation to anything else in our everyday lives, that we can be connected *online* to all kinds of audiovisual forms of intermedial music.

Thus, the music video has a long tradition and before the music video, there were other ways of combining music and visual media. After the so-called golden age of the music video (during the 1980s and the 1990s), new forms of music video production, new ways of experiencing music videos and new ways of *making use of* music videos have been established in the digital era (see Vernallis (2013) and Korsgaard (2017)). At twenty-first-century big record companies, professional directors and video producers are challenged by amateurs, fans and music lovers (on YouTube, for instance).

What makes the music video a music video is foremost the different combinations of both basic and qualified media involved, but also other qualifying media functions, such as the communication of ideas and values and being primarily a commercial *and* a work of art (at the same time). What is told, and what is sold? These are the questions that we should try to answer when we analyse specific music videos.

In the following section, the central perspectives on the medial and intermedial circumstances that define the music video in terms of audiovisual intermediality are presented so that you can continue your own analyses and explorations in this exciting qualified medium.

The music video must be understood in terms of intermediality. In at least two main ways it can be considered an intermedial kind of communication: 1) it is a (multimedial) combination of different basic media (sound and image), as well as of different qualified media (music and film); 2) it is also a media transformation, being an audiovisual result of the adaptation of a musical composition, in most cases a vocal music composition (Askander 2018, pp. 3–4). Here, the focus will be on the combination aspects.

A music video must include the basic media types of sound and image. It concerns the qualified medium of music being combined with moving images. If those two are not present, then there is no music video to talk about. But words – what about words? Actually, a music video can be produced to fit a non-vocal piece of music. So words are not necessary when defining a music video. But as we know, most of the music videos that have been produced do include words, in the oral performance of the singer in question. Very often, words are also shown on the screen, intra-diegetic words, as well as in the paratexts informing the viewers about the name of the performing artist, the title of the song, the name of the record company, etc. It is worth noting that most music videos that have been broadcasted (on MTV, for instance) do not include information on the director or the scriptwriter.

One of the initial and most crucial questions to raise when studying music videos is what are we looking for? Is it the music video as a whole, and the meaning produced in it? Or is the focus on the comparison between the music video and the corresponding piece of music, which has been released prior to the music video in question?

In this section, we have tried to map out the different kinds of media involved in a music video which enables us to discuss how these media forms work in the modes and modalities of such media. Here, we make use of a very well-known music video: Bob Dylan's *Subterranean Homesick Blues*.

Cards of poetry: Bob Dylan's Subterranean Homesick Blues

One of the most famous music videos ever was not even a music video per se from the beginning. Film director D.A. Pennebaker filmed and created *Subterranean Homesick Blues* to be the introductory part of his documentary *Don't Look Back* (*Subterranean Homesick Blues* 1967, one of the first 'rockumentaries' ever). In *Subterranean Homesick Blues*, we see a young Bob Dylan standing in a backstreet, with big paper cards in his hands. Words are written on the cards. When the song starts, Dylan starts throwing these cards to the ground, one by one. The words written on the cards are words from the lyrics of the song. Dylan doesn't mime the singing, and the viewer (and listener) don't see any musicians playing the music that can be heard on the soundtrack. The

combination of basic media is here quite obvious: mixed together are written words (on the cards), the sung and orally performed words, the moving images (though they look quite static) and the non-verbal but organized sounds of music (here there is 'only' the music, no other sound effects are used). Dylan is seen standing quite still, and what we primarily see is his small body movements and the cards repeatedly falling to the ground. Everything is shot in one take, adding to the feeling of the static, low-action situation in the clip. The background depicts urban scenery, which goes hand in hand with the rough, noisy rock music in the song. And regarding the words, the lyrics, we initially hear Dylan singing 'Johnny's in the basement, mixing up the medicine/I'm on the pavement, thinking about the government […]'.

It is not just the basic media (words, images, sounds) that are present. The clip can also be understood as a combination of different qualified media: poetry/literature, film and music. In terms of media modalities the Dylan and Pennebaker video discussed here is materialized through the *material* mode of a screen (cinema, TV or computer); that is, a two-dimensional, flat surface, and for the sound, through loudspeakers or earphones. *Sensorially* one can both see and hear the music video as it unfolds. Regarding the *spatiotemporal* dimensions, there are at least three important layers to take into account: the time and space of the watching and listening situation, the time and space of the lyrics' contents, and the time and space represented in the visual material.

What is being depicted visually might extend or narrow down what is going on and what is being represented in words in the lyrics of the song. In the Dylan video, we could say that the lyrics are communicating more time, and more things happening, than is the case in the imagery that we can see. The same goes for space, though the lyrics are also describing a wider perspective than the moving images do here. Finally, the *semiotic* modality should be understood in terms of iconic, indexical and symbolic signs. The moving images, the film clip and the sound recording of real events that actually took place: Dylan did stand in a street while being filmed; he and his musicians did play the instruments we hear. And the video contains symbolic, verbal language, which is written in the field of vision (on the cards, but also in the paratexts, which is the information initially shown in the clip about which artist and which song the clip is about, and so on), and can be heard in the orally performed song lyrics of the soundtrack.

Moving on from these observations on media, modes and modalities, one can interpret this music video in many ways. One of our main suggestions for an interpretation goes as follows. Bob Dylan wants to underline his status of being a poet. He denies the viewer the opportunity to see him (miming) singing and instead emphasizes both the sung and the written words. To that, one can add the fact that the famous American beat poet Allen Ginsberg (1926–1997) can be seen in the background during the whole clip, chatting with a musician in Dylan's band. Ginsberg's presence generates an aura of 'real' poetry because he is an established literary artist who belongs to the production of meaning, from whom Dylan borrows cultural capital.

What can be said about the more or less traditional music video, as we have tried to point out above, will not hold for all music videos in our modern digital mediascape (see Vernallis (2013) and Korsgaard (2017)). Next, we put forward a few modern examples of how music video aesthetics is practised in the twenty-first century. We begin with a digital interactive music video, and then move on to the so-called visual album.

New solutions: the music video in the twenty-first century

In 2010, the American indie rock band Arcade Fire collaborated with Google/Google Earth in the creation of the music video for the band's song 'The Wilderness Downtown' (2010). First, the viewer/listener/user is asked to type the name of the street and town or city where he or she was born. Then the video and the song start. The song is in a low key, with a kind of whispering singing style, and then a young man in a hoodie can be seen in the video, running through the streets of a city. Finally, the man stops and look up, and the viewer of the video sees what the person in the video is now looking at: supposedly, the house where the viewer was born.

This trick is achieved by using Google's project Google Earth and its GPS function; it works by using satellites and its aim is to create a photographic map of the whole world. The trick is also about making the viewer think, 'This video has been made especially for me'. By means of digital technology, the music video becomes personalized, almost intimately connected to the viewer and his or her childhood memories. Through the interactive force of this video, a highly personal experience seems to be made possible At the same time, some viewers may feel that the creators of the video are showing off a little, perhaps to deliberately make many viewers/listeners in the audience ask, 'Wow, how did they do *that*?'

During the 2010s, the concept of 'visual albums' gained more and more attention. Maybe 'visual album' should be replaced by something like 'audiovisual album' or 'intermedial album' to highlight the mix of media involved. Beyoncé (b. 1981) and her fellow directors created *Beyoncé* (2013), and later on *Lemonade* (2016) as visual albums. The first is more like a collection of music videos, without any obvious thematic connections. *Lemonade* is something else. Here we have a fully developed story (even though it is mostly a poetic story) that seamlessly knits together the songs and videos. In between the songs, we hear Beyoncé reciting poetry from poetry books by the Somali-British poet Warsan Shire (b. 1988). These passages are accompanied by visual material showing different settings from both urban everyday life and nature landscapes and sceneries. One could say that *Lemonade* is a collection of music videos, like *Beyoncé*, but here, the videos are not connected to each other in a random way. They are also 'glued' together by the intersections with the recitals of Shire's poetry.

So, *words* are spoken and sung, and visual text can be read in the final credits and in the title signs between the different 'chapters' in the film (like 'Denial', 'Anger', etc.). *Music* is played, and other *sounds* can be heard (both on-location

sounds and added sound effects); the moving *images* show the artist performing, old home video clips with Beyoncé's relatives, photos of people and depictions of different surroundings. All these basic and qualified media collaborate in producing meaning in *Lemonade*. Or, rather, they collaborate with the viewer/listener in producing meaning through the material, sensorial, spatiotemporal and semiotic modalities. Several stories and themes can be identified in *Lemonade*: through the lyrics one hears, through the actions and people in and the design of the visual material, and through the music (and other sounds), it becomes clear that this visual album is about race and gender in the history of the US (and maybe globally) but is also about the marriage and relationship of Beyoncé and Jay-Z.

Our suggestion following this investigation is that the best way of approaching music videos by using intermedial theory and concepts is to consider music videos in terms of having the following elements: 1) intermedial combinations of different sorts and 2) media transformations. When analysing music videos, a carefully mapped out description and structuring of qualified media types and modalities is important to understand the many different medial circumstances that are collaborating intermedially. It is just as important to then move on to interpreting and analysing the music videos in question with the aim of gaining an in-depth understanding of how the production of meaning occurs.

Conclusion

To sum up, in this chapter the ambition has been to shed light on intermedial combinations by discussing and analysing different qualified media types, such as comics, sounds in films, radio drama, singing and music videos. These all combine different basic media types but are also set in processes of intermedial transformations, for instance, the adaptation of a graphic novel into a movie, or a TV series that inspires a radio drama. In this chapter, however, we have focused only on the combination functions. In relation to all the combinations of basic media discussed here, we have strived to conduct analyses and to reach for the productions of meaning, which are found in between the media types in question; these productions of meaning collaborate with and/or oppose each other.

In one other aspect, the examples discussed in this chapter also challenge the reader/viewer/listener: one must have different kinds of literacy to really be able to identify certain references and to differentiate structures, forms and content to understand the media product in question. This is a specific challenge for intermediality analysis; if to analyse an intermedial act of communication, must not one, then, have the competence for all the media forms involved? The question is compelling, and should be paid more attention to in upcoming research on intermediality.

Notes

1 In addition to the TV series and the radio drama, the BBC produced a six-part spin-off called *The Body Farm* (2011). These belong to the same storyworld and could thus

be studied with reference to the concept of transmedia storytelling as defined by Henry Jenkins, that is, a representation of 'a process where integral elements of a fiction get dispersed systematically across [different] delivery channels for the purpose of creating a unified and coordinated experience' (Jenkins 2010, p. 944). However, this lies somewhat outside the scope of the present study.

2 'prequel, n'. In *Oxford English Dictionary, OED* Online. December 2020. Oxford University Press. http://www-oed-com.ludwig.lub.lu.se/view/Entry/150546?redirectedFrom=prequel& [Accessed 14 December 2020].

3 Information from the homepage of the BBC 4 radio drama.

4 The analyses in this section in general do not refer to any specific recordings or editions of musical and literary works; rather, they aim at any and all conceivable medial manifestations of a given work, including for example live performances of songs. Therefore, concrete references to works are in general not included.

Further reading: Sound design

Anderson, J.D. and Anderson, B.F. 2005. Part three. Acoustic events. In J.D. Anderson and B.F. Anderson, eds., *Moving image theory: Ecological considerations* (pp. 67–69). Carbondale: Southern Illinois University Press.

Eidsvik, C. 2005. Background tracks in recent cinema. In J.D. Anderson and B.F. Anderson, eds., *Moving image theory: Ecological considerations* (pp. 70–78). Carbondale: Southern Illinois University Press.

Isaza, M. 2009. Ben Burtt special: WALL-E – the definitive interview. *Designing Sound: Art and Technique of Sound Design*. http://designingsound.org/2009/09/ben-burtt-special-wall-e-the-definitive-interview/ [Accessed 1 January 2018].

Kulezic-Wilson, D. 2020. *Sound design is the new score: Theory, aesthetics, and erotics of the integrated soundtrack* (Oxford Music/Media Series). New York: Oxford University Press.

Langkjær, B. 2010. Making fictions sound real: on film sound, perceptual realism and genre. *MedieKultur: Journal of Media and Communication Research*, 26(48), pp. 5–17. doi:10.7146/mediekultur.v26i48.2115

Theme Ament, V. 2014. *The Foley grail: The art of performing sound for film, games, and animation*. Oxford: Taylor & Francis Group.

Thom, R. 2013. Notes on sound design in contemporary animated films. In J. Richardson, C. Gorbman and C. Vernallis, eds., *The Oxford handbook of new audiovisual aesthetics* (pp. 227–232). New York: Oxford University Press.

Whittington, W. 2012. The sonic playpen: Sound design and technology in Pixar's animated shorts. In T. Pinch and K. Bijsterveld, eds., *The Oxford handbook of sound studies* (pp. 367–368). New York: Oxford University Press.

References

Agawu, K. 1992. Theory and practice in the analysis of the nineteenth-century 'Lied'. *Music Analysis*, 11(1), pp. 3–36.

Allen, C. 2008. Radio narrative. In D. Herman, M. Jahn and M.-L. Ryan, eds., *Routledge encyclopedia of narrative theory* (pp. 481–482). London; New York: Routledge.

Askander, M. 2018. Immigration, paper planes, and a ka-ching! Understanding music videos through the concept of iconicity. *M&STE: Elektronisk tidskrift för konferensen Musik & samhälle*, 5(5), pp. 30–40.

Auslander, P. 2004. Performance analysis and popular music: a manifesto. *Contemporary Theatre Review*, 14(1), pp. 1–13.

Bateman, J. 2014. *Text and image: A critical introduction to the visual/verbal divide*. London: Routledge.

Bateman, J.A. and Wildfeuer, J. 2014. A multimodal discourse theory of visual narrative. *Journal of Pragmatics*, 74, pp. 180–208.

Bechdel, A. 2007. *Fun home: A family tragicomic*. Boston; New York: Houghton Mifflin Harcourt.

Bernaerts, L. 2016. Voice and sound in the anti-narrative radio play. In J. Mildorf and T. Kinzel, eds., *Audionarratology: Interfaces of sound and narrative* (pp. 133–148). Berlin; Boston: de Gruyter.

Bernaerts, L. 2017. Dialogue in audiophonic fiction. In J. Mildorf and B. Thomas, eds., *Dialogue across media* (pp. 205–223). Amsterdam, The Netherlands: John Benjamins Publishing.

Borčak, L.W. 2017. The sound of nonsense: on the function of nonsense words in pop songs. *Sound Effects*, 7(1), pp. 28–43. https://www.soundeffects.dk/article/view/97177/145962 [Accessed 16 November 2020].

Chignell, H. 2009. *Key concepts in radio studies*. London: SAGE.

Chute, H. 2008. Comics as literature? Reading graphic narrative. *PLMA*, 123(2), pp. 452–465.

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P.J. and Kuperberg, G.R. 2012. (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology*, 65(1), pp. 1–38.

Cook, N. 1998. *Analysing musical multimedia*. Oxford: Oxford University Press.

Frith, S. 1996. *Performing rites: On the value of popular music*. Oxford: Oxford University Press.

Gaver, W.W. 1993. What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5(1), pp. 1–29.

Grady, C. 2020. 'Space'. *Lunarbaboon*, January 18. http://www.lunarbaboon.com/comics/space-1.html [Accessed 15 December 2020].

Groensteen, T. 2007. *The system of comics*. Jackson: University Press of Mississippi.

Huwiler, E. 2016. A narratology of audio art: Telling stories by sound. In J. Mildorf and T. Kinzel, eds., *Audionarratology: Interfaces of sound and narrative*. Berlin; Boston: de Gruyter. pp. 99–115.

Jenkins, H. 2010. Transmedia storytelling and entertainment: An annotated syllabus. *Continuum*, 24(6), pp. 943–958.

Korsgaard, M.B. 2017. *Music video after MTV: Audiovisual studies, new media, and popular music*. London; New York: Routledge.

Kukkonen, K. 2013a. *Contemporary comics storytelling*. Lincoln: University of Nebraska Press.

Kukkonen, K. 2013b. *Studying comics and graphic novels*. West Sussex: John Wiley & Sons.

Kulezic-Wilson, D. 2020. *Sound design is the new score: Theory, aesthetics, and erotics of the integrated soundtrack* (Oxford Music/Media Series). New York: Oxford University Press.

Langkjær, B. 2010. Making fictions sound real: On film sound, perceptual realism and genre. *MedieKultur: Journal of Media and Communication Research*, 26(48), pp. 5–17.

Lutostanski, B. 2016. A narratology of radio drama: voice, perspective, space. In J. Mildorf and T. Kinzel, eds., *Audionarratology: interfaces of sound and narrative* (pp. 117–132). Berlin; Boston: de Gruyter.

Mader, D. 2007. The descriptive in audio-/radioliterature – a 'blind date'? In W. Wolf and W. Bernhart, eds., *Description in literature and other media* (pp. 179–213). Amsterdam: Rodopi.

Mandyczewski, E., ed. 1895. *Schuberts Werke, Serie XX: Sämtliche einstimmige Lieder und Gesänge*, 8(474), pp. 1823–1827. Leipzig: Breitkopf & Härtel.

Mildorf, J. and Kinzel, T. 2016. Audionarratology: Prolegomena to a research paradigm exploring sound and narrative. In J. Mildorf and T. Kinzel, eds., *Audionarratology: Interfaces of sound and narrative* (pp. 1–26). Berlin; Boston: de Gruyter.

Oxford English Dictionary, OED online. Oxford: Oxford University Press.

Rippl, G. and Etter, L. 2013. Intermediality, transmediality, and graphic narrative. From comic strips to graphic novels. Contributions to the theory and history of graphic narrative. In J. Fotis, M. Martínez and J.P.W. Schmid, eds., *Narratologia. Contributions to narrative theory* 37 (p. 191). Berlin: de Gruyter.

Schafer, R.M. 1994 (1977). *The soundscape: Our sonic environment and the tuning of the world*. Reprint. Rochester, VT: Destiny Books.

Spiegelman, A. 1991. *The complete MAUS*. New York: Pantheon.

Stein, D. 2015. Comics and graphic novels. In G. Rippl, ed., *Handbook of intermediality: Literature–image–sound–music*, vol 1 (pp. 420–438). Berlin: de Gruyter.

Theme Ament, V. 2014. *The Foley grail: The art of performing sound for film, games, and animation*. Oxford: Taylor & Francis Group.

Vernallis, C. 2013. Music video's second aesthetic?. In J. Richardson, C. Gorbman, and C. Vernallis. *The Oxford handbook of new audiovisual aesthetics*. Oxford; New York: Oxford University Press, pp. 437–465.

Wierzbicki, J. 2008. Shrieks, flutters, and vocal curtains: electronic sound/electronic music in Hitchcock's The Birds. *Music and the Moving Image*, 1(2), pp. 10–36.

Wilson Smith, M. 2007. *The total work of art: From Bayreuth to cyberspace*. London; New York: Routledge.

Wolf, W. 2016. How does absence become significant in literature and music? In W. Wolf and W. Bernhard, eds., *Silence and absence in literature and music* (pp. 5–22). Leiden; Boston: Brill/Rodopi.

Radio and TV referenced

The Body Farm. 2011. BBC, BBC 1.

The Unforgiven. 2018. BBC, BBC Radio 4.

Waking the Dead. 2000–11. BBC, BBC 1.

Films referenced

Atonement. 2007. Directed by Joe Wright. UK, France, USA: Universal Pictures, Working Title Films.

Corpse Bride. 2005. Directed by Tim Burton and Mike Johnson. USA, UK: Tim Burton/Laika Entertainment Production.

My Neighbor Totoro. 1988. Directed by Hayao Miyazaki. Japan: Studio Ghibli.

Pride and Prejudice. 2005. Directed by Joe Wright. France, UK, USA: Focus Features, Universal Pictures.

Showboat. 1929. Directed by Harry A. Pollard. USA: Universal Pictures.

The Birds. 1963. Directed by Alfred Hitchcock. USA: Alfred J. Hitchcock Productions.

Music videos referenced

Lemonade. 2016. Beyoncé. Directed by Jonas Åkerlund, Beyoncé, Kalihl Joseph, Melina Matsoukas, Dikayl Rimmasch, Mark Romanek and Todd Tourso. USA: Parkwood Entertainment and Columbia Records

Subterranean Homesick Blues. 1967. Bob Dylan. Directed by Donn Alan Pennebaker. USA: Pennebaker Hegedus Films.

The Wilderness Downtown. 2010. Arcade Fire. Directed by Chris Milk (Created in collaboration with Google).