INTO THE MULTIVERSE—STUDENTS COMMENTING ON AUDIOVISUAL SIMULTANEITY

VOLKER PIETSCH

University of Hildesheim

Abstract

In this article, I analyze the results of a qualitative study that investigated how a class of seventeen fourth graders (aged 10 to 11) from a primary school in Lower Saxony verbalize and construct the actions, feelings and perspectives of characters in the same shot by relating it to different soundtracks. The lesson was video- and audio recorded. The transcribed conversation was examined with qualitative content analysis. The results indicate the potential for Multimodal Literacy and Literary Learning that lies in the development of narrative interactions between image and sound in their simultaneity. They encourage examination of whether such approaches can effectively combine the dimensions of Frank Serafini's theoretical framework for Multimodal Literacy. They further suggest that the models of progressive literary competence acquisition still give too little consideration to the didactic potential of film in its media-specific multimodality.

Keywords: multimodality, (audiovisual) simultaneity, literary education, talking about film, soundtracks, story and character development, qualitative content analysis

Pietsch, V. (2024). Into the multiverse - students commenting on audiovisual simultaneity. L1-Educational Studies in Language and Literature, 24, 1-29. https://doi.org/10.21248/l1esll.2024.24.3.628

Corresponding author: Volker Pietsch, University of Hildesheim, Universitätsplatz 1, 31141 Hildesheim, Germany, email: pietsch@uni-hildesheim.de

1. INTRODUCTION

1.1 The multimodality of film in the discourse on literature and media education

Like other multimodal texts, a film is a "complex, multimodal entity […] utilizing a variety of cultural and semiotic resources to articulate, render, represent and communicate an array of concepts and information" (Serafini, 2014, pp. 12–13). Schmitz identifies five fundamental media modes as "spoken and written language, still and moving images, as well as audio (including music and sound)" (Schmitz, 2016, p. 331). For a long time, literature education was dominated by the primacy of the written word (see e.g., Serafini, 2015, p. 412, referring to Kress & van Leeuwen, 2001). However, the intermodal and intramodal relationships in texts are increasingly gaining attention of research (see e.g., Grabbe & Rupert-Kruse, 2013, pp. 16–17).

Serafini refers to the interactions of signs within the same mode as intramodal and the interactions of signs with signs in other modes as intermodal (see Serafini, 2015, p. 417). According to Serafini, understanding these interplays is one of the foundations of Multimodal Literacy (see ibid., p. 413). Furthermore, there are additional components of that to consider, one of which is particularly relevant for this article: the role of the recipients in constructing meaning potentials. Serafini summarizes this aspect in his three-part theoretical framework for multimodal literacy under the term "Structural Analytical Perspective" (ibid.). From a perceptual analytical perspective, the text is conceptualized as a multimodal object (in Serafini's example, as a visual object). An ideological analytical perspective conceptualizes text as a sociocultural artifact. From the structural analytical perspective, however, the text is considered a multimodal event.

> One's interpretation and modes of representation might change in different times and spaces based on different intentions, motivations, and understandings. […] The contexts and experiences of the reader are constantly changing, bringing different resources to bear on the semiotic event. (ibid., p. 417).

In the following article, we will focus on a short film that consists of only a single shot. Despite its formal reduction and lack of action, this film unfolds a complex form of intermodality.

The discussion will center on how a class of seventeen fourth graders from a primary school in Lower Saxony verbalize and construct the actions, feelings, and perspectives of characters in this same shot by relating it to different soundtracks. Thus, the focus shifts to how the meaning of the text is generated by the various combinations of visuals, audio tracks and the dynamic development of narratives by the students. As Serafini writes regarding picture books:

> How we come to view a particular image and how it is positioned in a composition affect the meaning potential of our interactions. […] By reconsidering text as a multimodal event, I am suggesting that readers have to pay attention not only to what is in a particular image or mode (intramodal considerations), but also how that image is positioned

> on the page, what is included nearby, and how it interacts with the other modes and features (intermodal considerations). (ibid., p. 417).

Films are texts that are structured to a particularly high degree by both intermodal and intercodal interactions (see e.g., Grabbe & Rupert-Kruse, 2015, p. 16). They usually utilize all five modes identified by Schmitz (see above) in various codes (moving images, for example, use cinematographic codes such as camera movement and editing, ibd., p. 20). Andrew Burn therefore specifically uses the term "metamodality" for film to do justice to the complexity of these interrelationships. "The prefix 'meta-' is used to indicade 'beyondness' and 'adjacency' —cultural forms and modes within, beyond and next to each other." (Burn, 2013, p. 5). Burn avoids using the term "codes" to refer to different sign systems within the modes. Instead, he distinguishes between "Orchestrating modes" (Filming and Editing) and "Contributory modes". The latter are embodied modes such as Dramatic action and Speech, Auditory modes such as Music, and Visual modes such as Lighting and Set design, each of which is further differentiated (e. g. ibid., pp. 5–9). This distinction can be helpful for analysis but can also lead to misunderstandings: For example, it is problematic to consider the "Orchestrating modes" as the medium-specific meta-modes of the moving image (Burn, 2013, p. 6), among other things, because aspects such as film acting, film music, or film production design differ in many respects from other forms of acting, music, or design and have undergone their own developments due to the medium (see e.g., Pietsch, 2018, pp. 59–61). The hierarchy between orchestrating modes and contributory modes established by Burn for text analysis (see e.g., Burn, 2013, p. 6), should also not lead to the assumption that there are corresponding hierarchies in terms of film conception or reception: Of course, for example, in film production, editing can adapt to the rhythm of the music; likewise, the attention of the audience can focus on the layering and interaction of characters and objects in the depth of the visual space (see e.g., Pietsch, 2018, pp. 99–101, 126–149). Burn himself rightly emphasizes: "Film Studies has arguably overemphasized filming and editing, at the expense of the other semiotic modes" (Burn 2013, p. 22). According to Burn, it is important to consider the "polyphonic structure" (ibid., p. 8) of the film, which he compares to the structure of a "fugue" (ibid., p. 8). On a more basic level than Burn's approach, the filmic representational field can also be described as bimodal—that is, audio-visual (see e.g., Grabbe & Rupert-Kruse, 2015, p. 21, referring to Anderson 1998, p. 80). "Image and sound take place within active reception in the form of resonance, that is, the acoustic dimension is able to serve as a receptive interpretative template for visual information." (Grabbe & Rupert-Kruse, 2015, p. 37).

Building on the above discussions, it is therefore highly relevant to examine the interplay between active reception and the various modes of film (especially its fundamental modes of sound and image). The question arises to what extent specific forms of meaning construction occur here, which may also have a particular potential to enhance students' understanding. Moreover, the self-reflection of the students regarding their achievements in constructing meaning should also be considered as an area of competence:

> Perception and interpretation are not seperate mental operations; rather they are thoroughly interconnected processes and any apporach to unterstanding […] multimodal texts must acknowledge this interconnection. (Serafini, 2015, p. 414).

In the academic German discourse on film and literature education, there is an emphasis on the narrative aspects of film: Some discussions present the various means of expression in film as an intriguing challenge from a didactic perspective as students need to discern and separate plot and character information from the simultaneous and diverse array of audiovisual codes that they perceive through their sensory channels (see e.g., Maiwald, 2005, p. 41). In contrast, a lot of these approaches do not explicitly address the importance of reflecting on one's own perception and (co-)construction of the film, as well as one's involvement in its narrative spaces, which may extend to other non-filmic constructions of reality. Thus, the structural analytical perspective and the ideological analytical perspective (see above) fall short. Consequently, the film seems to be implicitly treated as a finished and closed text that can be reconstructed and decoded retrospectively from a relatively neutral standpoint (see e.g. Maiwald & Kammerer, 2021, p. 38). In some cases, authors consider that students participate in the process of constructing the narrative by filling in the gaps or interpreting ambiguous scenes during their reception of the film, thus enhancing their imaginative abilities (see e.g., Abraham, 2009, p. 40). On the other hand, in models of literary competence that claim to be transmedially applicable, the unique aspects of film aesthetics are often given short shrift. These models of acquisition processes mostly fail to acknowledge that the construction of characters, plotlines, and mediating instances in film is multimodal and intercodal in both the production and reception process. As a result, they tend to overlook the complexities of film as a medium and its distinct narrative mechanisms (see e.g., Pissarek, 2018, pp. 135–158; Schilcher, 2018, pp. 199–228; Dürr, 2018, pp. 229–259). However, recent studies have been making media-specific differentiations based on these models, for instance, with respect to audiovisual texts, as in Scheubeck 2023, building on the model by Schilcher and Pissarek (2018).

According to the premises of this article, the multimodality of film has a unique potential to make students more aware of the constructed nature of fictional characters and their actions, as well as the recipients own involvement in these constructions. The simultaneity of various visual and auditory elements in film enables flexible separation and remixing of these components. The article aims to present the initial findings of a qualitative, explorative study that delves into how students explore a multiverse of potential stories, meanings, and effects of the same actions by associating them with different music and other sounds.

1.2 The interplay between sound and image in humanities and education

The following chapter will highlight the relationship between the research subject, humanities, and subject-specific didactics—a problematic situation, which shows the relevance of the study. Additionally, the chapter will summarize theoretical

foundations that will later be utilized for the expanded contextual analysis of the results (see Chapter 3.2). Last but not least, this summary of theoretical approaches aims to provide an initial introduction to the relationships between sound and image.

The term 'soundtrack' entails all the audible elements in a film, such as sounds, music, and spoken language. Unfortunately, the interrelationships between these different aspects, as well as between auditory and visual elements, are often overlooked in teaching practices due to subject-specific approaches (see Merlin, 2020, p. 3). Music lessons tend to have a longer tradition of exploring film music than do literature lessons (ibid.). In literature classes, the focus is primarily on dialogues (see Kammerer & Maiwald, 2021, p. 71). This becomes problematic in literature education because sounds and music are not merely supplemental to the visual aspect; they are inherently intertwined with it, particularly in children's and teen films (see Kurwinkel & Schmerheim, 2013, p. 125) as well as mainstream films (see Rabenalt, 2014, p. 14) and (post-)TV shows, which significantly influence film socialization (Barg, 2019).

In the context of analyzing children's and teen films, Kurwinkel and Schmerheim (see 2013, pp. 125–138) introduce the term "aurality" to describe the interweaving of sound and its significant role in guiding reception. Aurality is particularly relevant for the target audience of children up to the age of ten, as the sense of hearing plays a vital role during this developmental stage (see ibid., p. 95). While the empirical foundation provided by Kurwinkel and Schmerheim can be questioned, various other sources agree in attributing special significance to acoustic perception at an early stage: Even before birth, the fetus can experience an indirect environment through the sense of hearing (see ibid., p. 95, supported by Elsaesser & Hagener, 2011, pp. 179–180; Rabenalt, 2014, p. 69). In essence, music and other sounds play a significant role in shaping the reception experience of children and teens when engaging with a film, but not only for them. Drawing on neurophysiology (see Gorbman, 1987, pp. 63–69; Elsaesser & Hagener, 2011, pp. 179–180; Kurwikel & Schmerheim, 2013, p. 95; Rabenalt, 2014, pp. 104–106) and philosophical aesthetics (see Seel, 2013, pp. 43–46; Rabenalt, 2014, pp. 70–71), the aforementioned authors, along with others, categorize acoustic stimuli as more perceptually immediate compared to visual perceptions (summarizing the discourse: Pietsch, 2018, pp. 150–156). In their analyses, these authors (see e. g., Kurwinkel & Schmerheim, 2013, p. 134; Rabenalt, 2013, p. 106) focus on the emotional effects and bodily experiences related to sounds, music, and voices in film reception.

From a narratological perspective, these aspects are not only relevant for intensifying the perception of specific narrative moments, such as transgressions or climactic points. The synchronization of image and sound serves to strengthen and shape the audience's understanding of the narrative in multiple ways. Firstly, it establishes a connection between conventionalized expectations of the plot and certain effects. Additionally, it can imply classifications and evaluations of the characters or make them more ambiguous and open to interpretation. Moreover, it

influences the perspective through which the audience perceives the events unfolding on screen. In his discussions on his model of metamodal structures, Burn also notes, referencing van Leeuwen (1999), that music is not only "an emotional accompaniment to film, it has also narrative properties and […] carries particular social meanings" (e.g., Burn 2016, pp. 8–9). While speech is assigned to the embodied modes and music to the auditory modes in his hierarchy of modes, other sounds are not mentioned (see e.g., Burn 2013, p. 6, Burn 2016, p. 9). Since Burn does not fundamentally represent the distinction between the audio and visual layers of film in his structural model, he would have tomust assign sound to both modes—speech and music. This highlights the issue of assuming pre-filmic modes in contributory modes: The sound design of the film plays a different role than just recording sounds during shooting (and this has been the case even before digital technology).

Lastly, the combination of sound and image also plays a role in shaping the film's spatial dimension. According to Seel (see 2013, p. 38), music in film is significant in terms of temporality. It contributes to the film's continuous movement and utilizes techniques of time division. Both music and film direct the audience's attention to the moments before and after the current moment (see Seel, 2013, p. 40). Music, in particular, gives presence to elements that are not visually represented, thus expanding the cinematic space (see ibid., p. 47). This also applies to sounds and dialogue (Deleuze, 2015, pp. 291–295).

Markus Kuhn adopts Genette's terminology and applies it to film, distinguishing between "auricularization" (referring to the sound level) and "ocularization" (referring to the image level) (see Kuhn, 2011 pp. 127–194). This differentiation arises because information on the auditory and visual levels of a film is often differently focalized within the same shot (see ibid.). For instance, a whisper shown through gestures and facial expressions with the camera may remain unintelligible on the soundtrack, representing information shared among the characters but not with the audience (in Kuhn's terms, this is an example of external auricularization, Kuhn, 2011, p. 159). Moreover, the auditive and visual layers can simultaneously convey conflicting information. To create a coherent narrative, the abundance of simultaneous and mobile multimodal information in film must be managed (domesticated, so to speak). Filmmakers employ specific strategies to direct the audience's attention to certain aspects while diverting them from others (Pietsch, 2018, pp. 172–175). The way images, music, language, and sounds are combined in the film is carefully orchestrated to emphasize the information crucial for the narrative (see ibid., pp. 150–171).

However, even with this intentional direction, viewers can still perceive this supposed surplus on the periphery of the central narrative informations (Dästner, 2005, p. 93). These less emphasized details in both visuals and sound can also contribute to the audience's understanding of the narrative, particularly with repeated viewings and listening sessions. Multiple exposures allow viewers to notice and integrate these subtler elements into their comprehension of the overall story.

In their approach, Kammerer and Maiwald (see 2021, p. 70) consistently consider film music as functional, always subservient to the image, even in cases where music and image have formed a synthesis in the audience's memory. However, in certain instances, the rhythm of the montage, the content of the image, or the movement of the camera can synchronize with pre-existing composed music, as often observed in musicals and music videos. Even a monologue or dialogue can adopt lyrical qualities and align with the musical rhythm, without necessarily becoming a song (Pietsch, 2018, pp. 156–174).

Moreover, these techniques are not limited to children's and teen films; they are prevalent in genres aimed at adult audiences as well, such as thrillers and horror films. These genres frequently employ sequences designed to evoke either audiovisual or purely acoustic shocks, as seen in the jump-scare technique. The interplay of sound and image creates a powerful impact on the audience's emotions and senses. Various research works have attempted to categorize the different relationships between image and sound. Theorists often distinguish between an illustrative or paraphrasing image-sound relationship and a commenting or contrapuntal one (using different terminologies and with various ideological or ideology-critical dimensions) (see e.g., Balázs, 1961, p. 245; Pauli, 1981, p. 185; Eisenstein, Pudovkin & Alexandrov, 1998, p. 55; Adorno & Eisler, 2006, p. 15; Schneider, 2006, p. 89). When Burn illustrates his analytical model of metamodal structures using *Hamlet* (UK 1948, directed by Laurence Olivier), he also makes a corresponding assignment of music (see e.g., Burn 2013, p. 14).

However, one challenge is that these distinctions heavily rely on individual interpretations of the audience without taking this subjectivity into account. The concept of counterpoint implies that there is only one possible or correct interpretation of the given relationship between image and sound (see Dästner, 2005, pp. 83–91). The question of whether image and sound create a redundant relationship with each other also depends on the interpretation of the film's "statements" or messages, making it a matter of individual perspective. The physical involvement of the audience in the film experience is also overlooked by this concept.

Indeed, the idea of approximate redundancy between image and sound in film, where one could simply replace the other, is contradicted by Deleuze (see 2015, pp. 291–295). He emphasizes that the interplay of image and sound is not merely interchangeable. Deleuze introduces two distinct relationships between image and sound. Firstly, there is a concatenation between image and sound, where sound expands the image or refers to a variable whole, such as the past, the inner world of a character, or the perspective of a narrator (see ibid., pp. 291–295). On the other hand, Deleuze also describes an incommensurability between image and sound, particularly evident in radical manifestations of auteur films (see ibid., pp. 309–334).

Overall, Deleuze's perspective acknowledges the unique and dynamic role that sound plays in film, beyond mere redundancy or interchangeability with the image. He recognizes the power of the combination and the potential for cinematic innovation through the thoughtful interplay of these two elements.

Various categories are also applied in educational research:

Kurwinkel and Schmerheim (see 2013, pp. 130–131) distinguish between spatial (diegetic and non-diegetic, on-screen and off-screen) and temporal (synchronous and asynchronous) variations in their analysis of cinematic audio. However, they primarily view the temporal distance between sounds and images as relevant from a dramaturgical perspective.

Merlin (see 2020, p. 5), in contrast, proposes eight categories of cinematic audio analysis that he considers particularly suitable for l1 classes: auditive topography, symbolism vs. indexicality, underscoring vs. counterpointing, mood technique, leitmotifs, voice-off vs. voice-over, song text cues, and inter- and transmediality.

These categories, along with the additional research approaches outlined above, will remain pertinent for the broader contextual analysis of the second inquiry of the empirical data (see Chapter 3.2). In hindsight, it will become apparent that the results appear to correspond with certain recent research trends.

Nevertheless, these categories will not be used as a predefined framework for the analysis of the collected data. Instead, the film example, which is central to the study, will be presented with limited focus on technical characteristics, so as not to anticipate the conclusions made by the students. The aim is rather to maximize objectivity and allow the students' statements to guide the analysis. At this juncture, a noticeable shift in research can be observed. Here is an increasing focus on differentiating image-sound connections and recognizing the diverse contributions of recipients to these connections. Researchers are taking into account factors such as prior knowledge, perception, interpretation, and physical reactions, acknowledging the potential variability in how audiences engage with the audiovisual elements. Instead of rigidly defining hierarchies between image and sound, there is a tendency to view them as a cohesive unit, without absolute and unambiguous distinctions. This shift is also connected to the media-historical development of image-sound concepts in film production and the advancement of multi-channel sound systems in film distribution, which promise increasingly immersive experiences both in cinemas and home cinema setups.

Despite these developments, the simultanity of the visual and auditive levels is not yet prominently emphasized in literature education, either in comparing written literature to film or in the study of film itself. Although these tendencies are noticeable in the analysis criteria found in compendiums on children's and teen media or for teaching purposes, they have not yet significantly influenced the modeling of literary or film competence.

In summary, there is a growing recognition of the complexities in image-sound connections, as well as the dynamic role of audiences in shaping their experience. However, this understanding has not yet fully permeated literature education or the assessment of film competence, leaving room for further exploration and integration of multimodality in these fields. My research will therefore be guided by the following research questions: How do students connect different soundtracks to the same visual content when prompted by the teacher to create characters and storylines?

What reasons do they provide for their narrative constructions in relation to the various audiovisual combinations? To what extent do the findings indicate potential for Literary Learning and Multimodal Literacy?

2. METHOD

2.1 Data collection

In this qualitative study, I examine how a class of seventeen fourth graders from a German primary school in Lower Saxony verbalize and construct the actions, feelings and perspectives of characters in the same shot by relating it to different soundtracks. The lesson was video- and audio-recorded. I analyze the data using qualitative content analysis. The results indicate the potential for Multimodal Literacy and Literary Learning that lies in the development of narrative interactions between image and sound in their simultaneity.

The data was collected through a video- and audio recorded class discussion based on an audiovisual prompt. The data collection occurred in November 2018 by Loredana Seifert, who was a student at the University of Hildesheim at the time. This data collection took place during the first 45-minute lesson of a three-day project titled "Film music in elementary schools," which Seifert conducted as part of her master's thesis, completed in 2019. The conversations were transcribed by Loredana Seifert following Langer's guidelines (see 2013, pp. 519–523).

Seifert's objective was to introduce the students to the importance of film music in storytelling and its emotional impact in films. Additionally, she aimed to explore the development of imagination through film music as an aspect of Literary Learning in the teaching project (Loredana Seifert's research objective differed from mine). Prior to this project, films had only been used in the class as a means of conveying information, and they had not been the subject of study themselves. The students had previously engaged in focused listening during Integrated Studies for Primary Education ('Sachunterricht') in a unit on the senses. According to the teacher, the students had experience with tasks involving writing stories based on various non-film prompts in literature lessons. They had also worked on adapting poetry into other media, transforming poems into soundscapes and raps. In the initial lesson of the project, the students were presented with an audiovisual stimulus: a fifteen-second video by Angela Götz[1] showcasing the effect of film score ('score' refers to instrumental music specifically composed for a film). The video had three versions. In the first version, only non-music sounds were audible. The second version added an excerpt from a score by Alexandre Desplat, titled *Snape to Malfoy Manor*, taken from the soundtrack of the film *Harry Potter and the Deathly Hallows - Part 1* (2010). The third version included an excerpt from a score by John Williams, titled *A Window to*

1 https://www.youtube.com/watch?v=5w2eKr4X9EA

the Past, from the film *Harry Potter and the Prisoner of Azkaban* (2004). During this lesson, the students were not yet informed about the original sources of these music pieces. Initially, the teacher showed the film without any sound and then asked the students the following questions: What kind of character could that be? What might happen next and why? After the discussion, the same procedure was repeated with the three sound versions of the film. Finally, the teacher screened the different versions of the film again and invited the students to add their own interpretations and contributions.

2.2 The film material as basis for the class discussion

Given that all the details in the fifteen-second film can interact and potentially shape students' perceptions, a comprehensive description is provided below. However, it is essential to acknowledge that it is hardly possible to include every component of even such a short film due to its complexity.

The modes of film can be fundamentally divided into visual and auditory modes (for the reasons already stated, I deviate from Burn's framework, see Chapter 1.2). The visual modes include editing, camera, lighting and mise-en-scène, while the auditory modes include sound editing, speech, sounds, and music. As Burn states, all "need to decide at what level of granularity they want to work" (Burn 2013, p. 8); depending on this, even "a drop of glycerin" (Burn 2016, p. 8) on the actor’s forehead can contribute to meaning. For the film under consideration, particular attention to mise-en-scène is necessary: This includes modes such as costumes, set design, objects and actor (other possible modes such as make-Up and hairstyles or title design are not relevant in this example).

The fifteen-second clip consists of a single shot that appears to be set in a city (see figures 1–6).

The natural lighting is bright but not sunny. The camera maintains a calm and steady pace as it follows a character from behind, who is walking briskly along a path. The character occupies the foreground of the frame and is initially positioned slightly to the right of the center. The shot size is medium long, capturing the character from the ankles upward, and later, showing the character almost in full. The character wears a black, quilted parka with a slight waistline and the hood pulled over the head. A grey-white fur trim adorns the edges of the hood, adding to the overall appearance. The characters wears dark, long trousers and dark gloves. In the left hand, on a longer brown handle, carries a black bag (probably made of leather or imitation leather) with a brown top edge studded with several steel rivets. The character appears to be of rather slender build and its right arm swings back and forth as it walks. The camera shows the character for the first three seconds rather centrally from behind, then in a slight pan from behind to the right, then after nine seconds it switches to the character’s left side.

The path depicted in the film is sandy and wide, but it narrows in perspective on its right side, leading to a stretch paved with grey-blue stones. Along this stretch,

there are three bicycle racks and a green planted bed. The film clip ends just before the character reaches this narrowed portion of the road.

Figures 1-6 from left to right (screenshots)

In the first three seconds, as the character walks, they pass a manhole on their left. On either side of the path, houses can be seen. For the initial 13 seconds, a house with a grey and white striped ground floor and a passageway adorned with white columns is more prominently visible on the left side of the frame.

Topping the grey and white striped ground floor house, there are two more stories with rows of windows. Dark balconies can be observed at the distant corner of the building. In front of this house, there is a green area with shrubs, separated from the path by a fence on a low wall. At the beginning of the film, between the fence and the path, an area paved with grey-blue stones displays bicycle stands with bicycles and colorful children's bikes, a lamppost, and a young birch tree—supported and secured by a signposted wooden frame. Additionally, two white benches are visible at a further distance.

On the right side of the path, behind a much narrower paved area leading to the aforementioned narrowing, there is a lower building with a grey, plain roof. In the foreground on the left, a wide, open passageway or driveway is seen at the beginning of the film. This area is later separated by a yellow wall from a long passage of high glass exterior walls. On the right side of the path, before the passage narrows, there is only a grey lamppost visible.

In the direction the character is walking, the path is intersected by an asphalt street. Punctuated stone boundaries are present to prevent cars from turning onto the path from this street. To the left and right of the road, there are houses. The one on the left is white and has high, staggered windows. The house on the right is also white but features blue balconies in front of the windows. A road with parked cars at the edges runs between these houses until it is bordered by distant house facades on the horizon. Between seconds 1 and 3, a black bird, likely a corvid, flies up from the road in the left background and exits the frame on the left. From seconds 2 to 4, a red car drives from left to right across the road that intersects the path in the background of the picture.

Starting from the sixth second, another character comes into view on the left side of the foreground character. This new character is walking along the road from the left in the background of the picture and crosses the path. Initially, the character is seen behind the fenced boundary of the left front garden, but from the twelfth second, they are no longer hidden by it. A white car parked on the opposite side of the road enhances their figure.

The new character is wearing a dark cap, a grey top, a dark shoulder bag, dark long trousers, and light-colored shoes. They have a slender build and appear to be looking down. The character's light-colored face is shown in profile, but due to the distance, it lacks detail. As the clip ends, this character is at the left edge of the frame, walking to the right at approximately the same pace as the foreground character, who is shot at a slightly oblique angle from behind on the left and continues to walk straight ahead. In the first version of the film, only the crunching sounds of the character walking on the sandy path as well as traffic noise can be heard, at first attributable to the red car, then off-screen and increasingly louder as the character approaches the road.

The second version is, as mentioned, combined throughout with the beginning of *Snape to Malfoy Manor*, which can be found on the soundtrack album *Harry Potter and the Deathly Hallows, Pt. 1* as track 2. The piece starts with a muffled drumbeat accompanied by a repetitive motif played by cellos and basses throughout the excerpt, progressing from lower to higher notes with rapid and abrupt string strokes, only to subside again. In addition, a swelling trombone sound can be heard, the middle of which is accentuated by two short drumbeats. The long-drawn-out "Aah" of a choir, sung in bass, swells along with the wind instruments. The trombones subside and begin again with another drumbeat, this time in conjunction with horns. The string motif continues, now supported by clarinets and bassoons. From the fifth second of the film on, a lighter but equally repetitive string motif is heard, first played by violas, then also by violins, which progresses from lighter to lower notes and returns to lighter ones, at the same tempo as the first string motif and grounded by it. Meanwhile, the choir emits an increasingly bright 'aah' while trumpets now accompany the trombones and horns as their sounds continue to stretch and swell. At the climax of one of these increasingly loud and bright notes, the film ends, with the music fading out in the final second.

In the third version of the film, the image is combined with the piece *A Window to the Past*, which is track 7 on the soundtrack album *Harry Potter and the Prisoner of Azkaban*. In this film clip, the music starts playing from minute 02:31.

The motif in the music is played slowly in a minor key by a clarinet in the clarion range. It begins with a low E and alternates back and forth between higher, slower notes and slightly higher and slightly slower notes, encompassing a total of seven notes from a high E to a high G. It then returns to the high E twice and concludes on it the second time. Throughout these up and down movements, the sequence of notes is accompanied by four even longer, dark string chords.

The clarinet sequence is interrupted by a short pause, preceded first by descending and then ascending harp sounds. The pause, during which the strings continue to be heard, starts after the clarinet reaches high E for the second time and stops when it resumes higher, this time on a high A. From here, a sequence of notes unfolds again, ascending twice and descending again. The film ends abruptly with the second entry on a high A after such a pause, and the music quickly fades out.

3. CODING: THE CATEGORIES AS A RESULT OF THE PRIMARY EVALUATION

From the transcript, I inductively derived the following hierarchically organized system of categories (whereby various categories often overlap in students' comments, for instance, the anchor example for 1.2 also falls under category 1.1):

Designation	Definition	Example
1. individual components of the film	The utterance refers to individual components that the film contains visually or auditive.	I just wanted to say that I noticed that the man didn't have a hood on, but she the woman. (140)
1.1 Characters in the frame	The utterance refers to characters seen in the film.	There was a woman walking. (5)
1.2 Objects in the frame	The utterance refers to objects that can be seen in the film.	She also carries the (-) uh the (-) bag so far down. (49)
1.3 Sounds	The utterance refers to sounds that can be heard in the film.	Noise[2] and such. (60)
1.4 Music	The utterance refers to music that can be heard in the film.	it's just somehow so (-) so coarse-grained somehow (113)
1.5 Rhythm of the combination of sound and image	The utterance refers to the rhythm of the combination of sound and image.	*Eh? But she goes so fast. Somehow that doesn't fit.* (166)
1.6 Camera	The utterance refers to the camera that records the film.	The woman was filmed? (14)
2. Aspects beyond the film material	The utterance refers to something that points beyond the film material.	Um, maybe she just came from a hospital and has just, well, this music was so fitting for that she found out that she

[2] *The German word Rauschen can for example be attributed to wind or traffic noises.*

		somehow had cancer or something and now she's going home so very sad. (133)
2.1 Conditions of production and reception	The statement refers to the conditions of production and reception of the film.	(This is also always the case) in movies, so that it gets a little more exciting. (84)
2.1.1 Genre characteristics	In the utterance, elements of the film are identified as genre characteristics.	This sounds like 'to be continued' (with this MUSIC. This always comes during suspenseful moments) (108)
2.1.2 Potential or actual impact on the audience	The statement refers to the potential or actual effect of the film on the recipient.	The melody just makes it more exciting, too (-) because, for example, if there were *[our school song]* added, it wouldn't be that exciting. (82)
2.2 Motifs	The utterance refers to potential motifs of a potential film story.	that she found out that she somehow had cancer or something and now she's going home so very sad. (133)
2.2.1 Potential events in the film	The utterance refers to potential events in the film.	Maybe she is afraid that her husband will catch her, that sh- (115)
2.2.2 Potential future events	The utterance refers to potential events that lie in the future of the film events.	That right when she would go further there, that the man would go to the woman and maybe kidnap her. (179)
2.2.3 Potential past events	The utterance refers to potential events that lie in the past of the film events.	Maybe she also separated from her husband for example or something (164)

4. RESULTS

4.1 Examples from the secondary evaluation

In the secondary evaluation, a wider contextual analysis was conducted, which allowed for the inclusion of the entire understanding background of the interpreter for the explication (see Mayring, 2022, p. 67, 91). From this secondary analysis, I will now examine specific examples in more detail. I start with the translation of an excerpt from the transcribed remarks of a student about the version of the film combined exclusively with non-musical sounds. In the transcript, the students are numbered as S1, S2, etc., and the teacher as T.

> **S1**: But it could also be that the man walks toward the woman and the woman goes a bit diagonally because she is perhaps afraid? Of the (--) man?

This extract is remarkable because it is the first time a student brings into play the character that can be seen distantly in the background of the picture. This does not mean that this character was only noticed in the version with sounds, but it is the first time they are named as relevant to a possible motivation of a character and a (still rudimentary) potential story. S1 does not explicitly make the connection

between their reflection and the effect of the sounds, but 44 seconds earlier, they had already reacted to the screening of this version when the clip was over:

> **S1**: Totally creepy ((winces))

After the earlier screening of the film without sound, however, S1 had expressed another suggestion for the plot motivation of the character in the foreground:

> **S1**: She's going shopping.

In this regard, S1's new proposal aligns with her effect-related utterance regarding the noise version. The creepy effect of this version is consistent with the reception of early sound films in the late 1920s and early 1930s. In the transition from silent films, which were typically accompanied by live music during screenings, to early sound films, dialogue, sounds, and music were not always presented simultaneously. This was partly due to concerns about overwhelming the audience with auditory stimuli and technical limitations in recording and mixing sound elements separately. As a result, the extensive absence of music in certain early sound films created an unsettling and uncanny effect, especially when combined with specific subjects or scenes (see Kracauer, 1985, p. 187; Oates, 1991. p. 498–510; Spadoni, 2007, p. 78; Pietsch, 2018, p. 161–162). As Burn points out, concerning this era of film history: „We sometimes think of multimodal texts in forms of richness and replete semiotic texture; but lack, gaps, silences and 'workarounds' are just as important." (Burn 2013, p. 4).

S2 associates the loud stepping noises of the character in the foreground of the picture with a specific type of shoe, which, interestingly, does not match the shoes that are actually visible in the film.

> **S2**: Um—her footsteps like so. The steps like that. ((imitates the staccato sound under the table)). She was wearing heels!

The character in the foreground is quickly categorized by the students as being female, despite being seen from behind and having only a few stereotypical gender attributes (such as the bag and slightly waisted parka). On the other hand, there are few indicators of the gender attribution of the character in the background, except for the absence of clichéd feminine connotations. Some students classify the loud footsteps on the sand and the traffic noise as ordinary sounds. A few others interpret them as wind or rain..

> **S3**: I would say (-) everyday(-)noises.

Despite this, the students sketched out typical everyday actions of the character in the foreground exclusively after watching the previous completely soundless version (such as shopping, going home from work). Some students initially perceived this version as lacking a spectacular, conflict-ridden story. This reaction was evident in a disappointed and sarcastic response immediately after viewing.

> **S2**: W o w. (--) what a blast.

After the noise version, the students shift their narrative interpretation towards a crime story based on their statements. Some students associate the traffic noise with inclement weather. This and the isolated footsteps align well with conventional threat scenarios in crime stories or thrillers. However, these connections between the auditive elements and potential effects on reception or genre characteristics are not explicitly made by S1 or other students who build upon S1's scenario. In contrast, the film version that includes the music piece *Snape to Malfoy Manor* appears to evoke a more distinct reaction.

> **S4**: So, yes, well, um. For example, she comes like from work and this suspenseful music is like because there comes another person and no clue
>
> **T**: These two people, could they meet?
>
> **S4**: Sure.

S4 explains the function of the music by its reference to this approaching character. They choose the term "suspenseful" for it, which conventionally implies the anticipation of an event by the audience. At this point, the teacher's remark directs them to this event.

> **S2**: Um maybe (--) um maybe somebody just comes someh- there she is coming from home, has a date with this strange guy and um um sounds almost a bit like they are about to beat each other up.
>
> **S5**: That's right.
>
> **T**: Pick somebody.
>
> **S2**: (---) S1.
>
> **S1**: This sounds like ‘to be continued’ (with this MUSIC. This always comes during suspenseful moments) S6.
>
> **S6**: It could also be that because she's got a bag and she's walking pretty fast, that she has stolen something or something like that.

In the film version with music (*Snape to Malfoy Manor*), S2 associates the music with an impending escalation of violence. S1, on the other hand, abstracts from specific events in the film and identifies the combination of music and image as a genre convention, particularly typical for cliffhangers in series and generally for suspenseful situations. S1 likens the music to the producers' verbal announcement to the audience, "To be continued....". S6 introduces new observations of visual details, such as the bag and the speed of the character in the foreground. In this version, students, including S6 and S4 before, ascribe meanings to objects and characters in the film. The music is presented as a soundtrack overlaid on the film and does not emanate from specific sound sources anywhere within the image. In this process, the students reorganize the image, attributing meaning to previously neutral details that now motivate the visible and potential progression of the film's story. The term "suspenseful" used by the students to describe the music aligns with Hitchcock's preferred

technique of using suspense to evoke emotions in the audience, even for actions that may seem mundane to the characters in the film (Hitchcock as cited in Truffaut, 2009, pp. 62–64). The bag, the fast walk, and the character in the background were initially neutral elements, but they now become "spots in the field of vision," as described by Dolar (2002, p. 130), due to the surplus knowledge of the film viewers, or listeners in this case, who perceive the music as off-screen. According to Dolar's interpretation of Hitchcock, this surplus knowledge also leads to a lack of knowledge, as the audience is confronted with uncertainty about the exact outcome of the anticipated event or the true meaning of the objects or characters. The 'surplus of information thus turns into a deficit' [ibid.] or lack.

The following excerpt highlights this information gap between the characters in the plot and the audience:

> **T**: That would also make a good story. Would the story that she stole something also fit the music we just heard?
>
> **Reactions**: Yup / I don't think so / Not really
>
> **T**: S2's saying: "Not really." Can you try to describe the music and why do you think that?
>
> **S2**: Well, I don't think so, because um (-) well, it's like (-) it's like somehow (-) somehow
>
> coarse-grained so that (-) actually everyone would listen up. And it should (-) you shouldn't really know that the person has something planned. And it's[3] somehow already so (--)evil but I don't think that really fits

The statement made by S2 is not entirely clear regarding which narrative level they are referring to. When S2 says, "so that actually everyone would hear," they could be assuming that the music is intradiegetic and the characters in the film can also hear it. This would be atypical for this scene, considering typical cinematic conventions. Alternatively, S2 might be assuming that the music is extradiegetic. In this case, S2 may feel that the music reveals too much to the audience, potentially spoiling the suspense or mystery of the scene. This interpretation of S2's statement aligns more with the insights, average children of this age may have into conventional filmmaking practices based on regular experiences (Medienpädagogischer Forschungsverbund, 2018).

Regardless of the previous question, S2 finds the music track aesthetically unsatisfactory for a secret theft. They use the term "somehow coarse-grained" for the music, which is an interesting choice since this term is commonly associated with visual impressions (such as "grainy" pictures). It raises the curiosity of what piece of music S2 would prefer for this scene if given a selection of genre-typical melodies. For instance, they might choose one of the upbeat, jazzy themes often used for characters like thieves in subgenres such as caper or heist movies, where sophisticated

---

[3] [Or should it read "she"? Does S2 refer to the music or the 'woman'? In German, the grammatical gender of 'music ' is female.]

and elegant crimes are involved (e.g., *The Pink Panther*, UK / US 1963, directed by Blake Edwards; *Ocean's Eleven*, US 2001, directed by Steven Soderbergh). However, in order to not resorting to speculations, I continue with the transcript analysis.

> **T**: Uh-hu. Do we have any other ideas besides theft. S7?
>
> **S7**: Well, I wouldn't have any other idea now, I just wanted to say: Um, so, it could fit the music after all because, um, she's running so fast. Maybe she is afraid that her husband will catch her, that sh...
>
> **S8**: In that case, if it were me, I would stop

S7 does not immediately create a new story but focuses on the fit of the music. For the first time, a student notes that the speed of the character in the foreground is not only related to motivation but also to the rhythmic interplay of the movements with the music. However, S8 raises a logical objection, not to the consideration of the rhythmic interplay, but to S7's subsequent narrative variant, drawing a comparison with S8's own potential behavior in a situation like that.

> **T**: Okay, S5's got another idea. Shoot. S5.
>
> **S5**: That the woman has quite a lot of money in the bag and is bribed by this man. And (-) if she alerts the police that he's doing something to her.
>
> **T**: So S8 wasn't paying attention back in the corner just now. Repeat what you just said, really loud.
>
> **S5**: SO. Um, the woman could have quite a lot of money in the bag and, um, the man bribes her. If she doesn't give him money, (-) something will happen and if she alerts the police, (--) she will die

In this version of the film, the students create stories that clearly align with common motifs found in crime and thriller genres. Even if some students may be familiar with some of these motifs from documentary formats or their own experiences, students like S1 and S2 explicitly voice their expectations related to certain fictional genres. Those reactions differ from those to the following film version combined with the track *A Window to the Past*:

> **S9**: Did she come from the cemetery? S2.
>
> **S2**: Um, maybe that she, um, just, um, (-) for, (-) um, well. For example, she was on holiday and, um, then she was in Italy, for example, and, um, during that time her dad died at home or something and now she's just going to the funeral. (--) S1.
>
> **S1**: It just looks like a funeral because she's dressed all in black. (--) There at the cemetery you kneel down and grieve and stuff because she's dressed all in black. S5.
>
> **S5**: Yes, she comes from the cemetery and was just at her mother's grave and put candles there that she had in the bag.
>
> **T**: S2 now gets the last word, and then we move on
>
> **S2**: So, it could just be, for example, that now, somehow, um, I'd say, her husband has left her

The stories developed by the students for this version of the film could draw inspiration from fictional narratives; however, they could also be inspired by personal experiences, or even worries derived from real life. Interestingly, the children do not explicitly mention genre conventions as such, which may be because music commonly perceived as sad is also encountered in everyday life outside of fictional narratives.

On the other hand, suspenseful music is typically associated with films, radio plays, computer games, or theater productions. It is less commonly found in other contexts like religious music or pop music, where pieces with connotations like "aggressive" or "haunting" may be more prevalent (for example in music genres like Death Metal or Dark Wave).

Once again, the students attribute special emotional value to the bag (respectively to the imagined objects inside), and they also consider new visual details, such as the color of clothing. The color black is linked to the motif of mourning, which matches the mood of the music in this variation.

Furthermore, a student perceives a characteristic of this combination that goes beyond the subject matter and also relates to the rhythm. When the teacher shows this version again, during the screening S2 reacts as follows:

> **S2**: Say what? But she's walking so fast. Somehow that doesn't fit.

S2's dissatisfaction with the slow music in relation to the fast walk could be influenced by the fact that the sounds of the footsteps are also audible and may interrupt the elegiac melody like a beat. In any case, during repeated viewings of the film, S2 perceives this combination as aesthetically unsatisfactory. This mismatch between the music and the image is quite unusual in film, where rhythmically matching music and image is a common practice.

This situation brings to mind the term "aurality" coined by Kurwinkel and Schmerheim (see 1.2), which refers to the interplay of sound and image that guides reception, especially in films aimed at children. In children's films, the soundtrack goes beyond being a mere complement which contributes to the suggestion of a self-contained world (see Kurwinkel & Schmerheim, 2013, p. 98). Instead it creates an immersive film experience through immediate bodily reactions (see ibid., p. 98) when the rhythms of movements, music, and montage are interconnected.

Kurwinkel and Schmerheim attribute the importance of the auditory channel for children to biology and developmental psychology. However, S2's reaction does not necessarily indicate limited experience with cinematic complexity due to an exclusive exposure to children's films. Indeed, when discussing contrapuntal film music in contrast to illustrative film music, prime examples often come from films aimed at adults and created in modern or postmodern styles (see Schneider, 2006, pp. 89–90; Rabenalt, 2014, pp. 167–176). However, it is essential to note that even in these cases, the contrasts between music and image typically refer to content dimensions, which can be subjective and debatable. As previously mentioned, the contrast

between contrapuntal and illustrative music is problematic and not always straightforward. In *A Clockwork Orange (*UK / US 1971, directed by Stanley Kubrick), for example, gangs beat each other up set to the melodies of Rossini. This example can similarly be regarded as contrapuntal, thereby contrasting the atavistic behaviour with cultural zeniths. Nevertheless the enjoyment of violence on the part of the offenders might also be viewed as an anthropological constant, harmonizing with the appreciation of music (or potentially intensifying the violent experience). However, it's worth noting that in this and other examples, sound and film image, as well as sound and film editing, still harmonize regarding their tempo. So, S2's critical remark about the lack of rhythmic match between the music and the fast-walking character would not be less likely if she had seen many of the canonical auteur films by Kubrick and others.

Interestingly, upon closer examination, S2 revises her own previously sketched story based on the formal aesthetic quality of the film when re-watching and re-listening to it.

4.2 Connections between categories and the different versions of the film clip

To summarize the foregoing, in the completely soundless variant of the film, the focus is mainly on the character in the foreground (category 1.1), and the only object mentioned is the bag (category 1.2). However, as soon as sound is introduced in the subsequent variants, the students start to relate the character in the background (category 1.1) to various motifs (category 2.2) suggested by the sounds.
Furthermore, in the variants with sound, other objects such as the characters' clothes, the floor, or vehicles are introduced into the students' utterances, and they start to attribute possible sources to the sounds they hear (category 1.3). Additionally, there is a mention of a supposed source of sounds explained by its potential causality for the presence of an object.

> **S9**: Um (--) in any case, you can hear that it's also windy, because that- (-) (maybe that's why) she pulled on the hood.

Indeed, to achieve a deeper level of analysis, it would be possible to further differentiate subcategories within 1.1 and 1.2, focusing on individual film-aesthetic components such as gesture or color. However, given the available material and the intended clarity of the category system (see Heins, 2018, p. 305), it may not be proportionate to undertake such a level of differentiation.

In the "suspense music" variant, color becomes relevant for the students' interpretations, suggesting that the music has a different impact on how they perceive visual elements. However, objects (category 1.2) that are neither carried by the characters nor considered as sound sources are not discussed. For example, buildings, plants, or the bird in the scene are not mentioned in relation to the sounds heard.
The buildings and their surroundings in the film provide some interesting details. On one hand, the area appears to be a well-maintained and moderately nature-

integrated residential neighborhood. On the other hand, the buildings themselves have a sober and somewhat bleak appearance, with empty windows and balconies. However, these surroundings do not seem to provide any specific impulses or motifs (category 2.2) beyond the everyday actions expressed when watching the soundless film.

Moreover, parts of the buildings that could potentially be relevant to the narrative, such as the open passageway or driveway on the right, or the closed windows and balconies (so, all the places where someone could suddenly appear or disappear), are not prominently featured in the film. Instead, they are either briefly shown or kept at a distance behind bushes and a fence.

Overall, the film's surroundings appear neutral enough not to significantly influence the students' narrative constructions or interpretations beyond the immediate actions of the characters observed in the soundless film. The surroundings in the film also do not hinder the development of crime and melodrama motifs in the other film versions. The sounds (category 1.3) appear to provide impulses for the students to draw on genre characteristics (category 2.1.1) in developing their stories and understanding the motivations of the characters. Additionally, the sounds seem to have other atmospheric effects on the recipients (category 2.1.2). However, unlike with the music (category 1.4), these connections with the sounds are not explicitly mentioned in the students' statements.

This can be observed in the contrast between the first articulated reaction to the "sounds"-variant (e.g., S1's "Totally creepy") and the first articulated reaction after the presentation of the "suspense music"-variant:

> **S.10**: Dude this m u s i c.

Indeed, it is only after the introduction of the music variations that students relate to the production and reception conditions of the film (category 2.1), primarily in relation to the "suspense music"-variant. Unlike with the sounds (category 1.3), the students do not address the music (category 1.4) in a way that would stand alone without overlapping with another category. There are no isolated references to specific instruments or technical terms used to describe the music.

Instead, the students' language about the music is always related to something else, directing attention to specific elements within the film or allowing certain aspects to emerge and be interpreted. In the soundless variant, motifs (category 2.2) exclusively revolve around everyday actions, involving only characters (category 1.1) and objects (category 1.2) in the frame. On the other hand, the "sound"-variant, "suspense music"-variant, and "sad music"-variant give rise to the development of dramatic motifs. The dramatic motifs in the "suspense music"-variant are likely influenced by experiences with fictional genres (category 2.1.1), but the same cannot be said with certainty for the "sad music"-variant.

The camera (category 1.6) is mentioned only once, and it is unclear whether it is considered intradiegetic (this would mean that the students assume that the camera is a part of the narrative, operated or installed by another fictional character) or

extradiegetic. References to editing have not been made at all, although it would be interesting to analyze a film excerpt with more than a single shot and dynamics of the camera work that are more independent of the characters' movements. However, one student's attempt to assign the film to the genre of GIF (category 2.1.1) is related to its brevity and the repeated visual level, resembling a montage effect, even though the repetition in this film is not immediate, unlike in a GIF.

> **S.11**: Maybe it's a GIF
>
> **S.2**: A what?
>
> **S.11**: A GIF.
>
> **S.5**: What's that?
>
> **S.11**: A GIF is (-) that is, for example, when I do like this ((raises the hand)) and (that is then always repeated) ((repeats the hand movement, as in a loop)).

The students express themselves through gestures only if reacting to sounds and music, and they do so repeatedly. They wince noticeably, imitate the loud footsteps with their own feet, and hold onto themselves as if they were afraid. This is probably not because they lack the words to express themselves, as they have a diverse vocabulary of effect-related adjectives at their disposal (such as exciting, gruesome, creepy, cruel, angry, irate, coarse-grained, clumsy, big-headed, sad, alone, abandoned). Instead, their use of gestures seems to be an attempt to convey the intensity of their reception experience and the potential effect on other viewers more effectively than words alone can do. The students' references to future events (cat. 2.2.2) occur more frequently in relation to the "suspense music"-variant, while references to past events (cat. 2.2.3) are more often made in connection with the "sad music"-variant. This observation aligns with the thematic context in which the music pieces were composed. *Snape to Malfoy Manor* (the "suspense music"-variant) already hints at movement towards a specific location in its title and increases in intensity towards a climax, evident in the stretched tones and the repetitive, fast, and forceful motif. On the other hand, the excerpt from *Window to the Past* (the "sad music"-variant) is in line with its title through the recurring sequence of notes, slow and contemplative with inner slowdowns. While there is some development as the sequence continues higher after a pause, there is no clear build-up towards a significant climax in volume, pitch, or orchestration richness. During the "sad music"-variant, S2 notices the lack of rhythmic synchronicity (Cat. 1.5) between the music and the continuously dynamic movements and step sounds of the character in the foreground. This realization prompts S2 to reevaluate and correct their own story designs (Cat. 2.2) based on this discrepancy between the audio and visual elements.

5. DISCUSSION

5.1 Limitations

This is only an exploratory study, which, due to the small number of participating students, cannot aim for generalizations. It only allows for making initial assumptions and raising interesting questions. The classroom discussion is a collective process that, in this particular configuration and dynamics, will be highly individual.

> [S]tudents may draw upon other readers' ideas to construct and revise their own interpretations. In addition the teacher may unknowingly privilege or marginalize particular interpretations during a class discussion. (Serafini, 2015, p. 418).

The order in which the film variants were shown also has an influence, which could not be contextualized here due to a lack of sufficiently large samples with alternative sequences.

It remains open whether the students discovered certain characters and objects through sound and music. Nor can it be said whether they developed their stories solely or primarily through auditory stimuli. However, what can be discerned is that and how the students use sound and music in a variety of ways to justify their ideas and conclusions. Where the students got some of their further associations and inspirations for plot points also remains in the realm of considerations that can be more or less plausibly but not confirmed. Further knowledge about the students' social background would be needed here, especially regarding their media consumption.

Another limitation lies in the specific nature of the film clips and the pieces of music used. Differences in the various modes could bring significant changes in what stimulates the students for their stories: In particular, less ambiguous images paired with less thematic distinct pieces of music could give the image a greater dominance in the relationship to sound.

Additionally, it cannot be determined whether the film clips merely reveal the students' knowledge (more or less explicitly) or if the setting has the potential to promote developmental and understanding processes. A long-term study would be necessary to address this.

5.2 Implications for literature education

The students draw upon the diverse and simultaneous aesthetic elements of the film in its multimodality, intertwining them to construct different stories with reasoned justifications. Some students thus demonstrate a certain flexibility in focusing on the literal or denotative contents in a multimodal text, as captured by Serafini as part of Multimodal Literacy under a "Perceptual Analytical Perspective" (e.g., Serafini, 2015, p. 4, see also Chapter 1.1).

This flexibility is evident in that the children are confronted with a text whose intermodality is in flux, or rather with four similar multimodal texts in quick succession.

The audio layer does not primarily complement the impressions conveyed by the images; rather, the images are significantly imbued with meaning through the sounds and music (S12: "The music always makes um a difference."). Moreover, the sounds and music direct attention to various elements on the visual level. Regarding the music, students articulate their knowledge of production and reception circumstances, thus explicitly addressing the construct character of films. The manipulation through the combination of different channels is recurrently thematized in a media- and self-reflective manner. From here, economic, social, and ideological conditions (e.g., the cliffhangers mentioned by the students themselves to engage the audience with the product, or the implicit gender clichés picked up by the students in relation to certain typical genre scenarios or other societal conventions) could also be accessible topics for discussion. When the students quickly assign genders to the characters, it becomes evident how, in line with Serafini's perspective, the Ideological Analytical Dimension constitutes a relevant part of Multimodal Literacy (see Chapter 1.1): "How various cultures, ethnicities, races, and genders are represented in multimodal texts is an important consideration […]. Teaching students to recognize and question stereotypical depictions through written language or visual images, helping them to rethink how certain people are represented […] are foundational to this endeavor." (Serafini, 2015, p. 419, referring to Lehr, 2008, and Harste et al., 2000).

However, the awareness that the students are here constructing to a large extent instead of reconstructing, and that they are not only entangled in the cinematic space but also co-designing it, can easily be promoted or made explicit on the basis of the alternative variants they propose. Therefore, the study also highlights the potential to promote Multimodal Literacy from a Structural Analytical Perspective (e.g., Serafini, 2015, p. 413, see also Chapter 1.1).

Although the musical pieces are themes in the tradition of the classical Hollywood style, which in turn draws heavily on late Romanticism, employing leitmotifs (see Rabenalt, 2014, pp. 147–166, Moormann, 2009, p. 12), while also incorporating influences of new, atonal music (see Moormann, 2009, p. 15), they still prove to be open enough in combination with different images and different students to evoke a variety of descriptive terms and motifs and to develop actions. Thus, Seel's rather clichéd statement that the literary text is like a score that readers bring to performance and where they have to supplement the narrated story with their own imagination, while the film is a presentation that binds its audience more tightly to its own process (see Seel 2013, pp. 120–121) can be contradicted.

According to their expressions and gestures, the students do indeed perceive sounds and music in a physically intense way. However, in the case of this example, the images also lack action, do not contain relatively spectacular characters or objects, and do not show any formal fractures or striking accents through camera and editing. What can be said based on the results is that the physical and emotional

perceptions do not seem to hinder a well-justified reference to the various signs in the film and the development of a contingent story, but rather set it in motion and motivate it. Gestures are used several times by the students to complement the verbal language, but not at the expense of it. Thus, the construction of the film's story based on multimodal reception is also a physically holistic activity, which in turn is partly multimodally, namely audio visually mediated. So, these findings indeed seem to align with recent trends in research, encouraging a fresh perspective on both visual and auditory elements (by de-emphasizing hierarchy in their interaction), as well as the reception and construction of films (by considering the diverse prior knowledge, perception, interpretation, and physical reactions of the film viewers and listeners).

When it comes to modeling Literary Learning, the analysis of the students' responses further support the urgency of considering both the multimodality and the influence of recipients on the construction of meaning in literary contexts.

The analysis of the students's responses indicates that considering the simultaneity of sound and image, they explicitly articulate their knowledge in various dimensions of Literary Learning. This observation could be further explored in future investigations. These aspects are interconnected and can be linked to concepts proposed by Spinner (2006). Notably, the development of imagination, understanding narrative and dramaturgical logic, conscious engagement with fiction, and familiarity with prototypical genre characteristics seem to be key points of connection.

Central to all of this is an active awareness of language composition, not just in its narrow linguistic sense, but also of "language" referring to artistic style and as a metaphor for communication in different media. In the context of the film, we can also extend this notion to encompass "cinematic" or media-specific composition. Moreover, the transcript highlights that some students are capable of reflecting on their subjective involvement in the film and adjusting their initial impressions based on accurate perception. There are indications in the students' contributions that the medium of film, especially regarding the relationship between image and film music, can initiate levels of competence acquisition at an early stage, which are considered to be the most advanced in various models: For instance, the ability to grasp characters as constructs and representations is placed at the highest, fourth level of learning progression in the dimension "recognizing and interpreting features of the character" in the model of literary competence by Schilcher and Pissarek (2018, p. 324). Similarly, according to Spinner, understanding the perspectives of literary characters, bringing into play the narrative style and the perspective of the narrator, and relating this to the characters' points of view, their relationship constellation, and their world is particularly challenging (see Spinner, 2006, p. 10). The film, with its multimodality, is particularly suitable for providing access to such seemingly higher levels of abstraction, especially as it can draw on early-established experiences from children's media socialization compared to written literature (Pietsch, 2022, pp. 203–204). Films could be highly beneficial in inclusive education because they present characters and objects in a visible and audible manner. Students can identify

and verbally describe these elements together without the need to accurately decode specific words or sentences in a text. Moreover, students can draw from their diverse media experiences, which are often more varied or frequent in the medium of film compared to writing-based media. With regard to the main competence fields of Kammerer's and Maiwald's model of film didactics (see Kammerer & Maiwald, 2021, p. 39), it becomes apparent that not only film as a symbol system, film as an action system, and the transitional area between the two can be linked with the help of perceptions of the simultaneity of film. Also, the reflection of one's own film reception as a further significant competence is an inherent part of this linking performance.

Serafini's comprehensive framework of Multimodal Literacy underscores that the perceptual, structural, and ideological dimensions must be interconnected to achieve goals such as those of Literary Learning and understanding. Although this is merely an exploratory study, it nevertheless allows for initial encouraging assumptions that warrant further investigation. The film's multimodality allows for the selective amplification or attenuation of specific channels of perception, thereby enhancing or altering the focus on the visual and auditory aspects. This interplay between form and content, along with the ambiguity of individual elements, becomes perceptible and tangible through the film's audiovisual montage technique, which can be employed by teachers.

REFERENCES

Abraham, U. (2009). Filme im Deutschunterricht [*Films in German language teaching*]. Seelze-Velbert: Kallmeyer in Verbindung mit Klett.

Adorno, T. W. & Eisler, H. (2006): Komposition für den Film [*Composition for the films*]. Frankfurt/Main: Suhrkamp Verlag.

Anderson, J. D. (1998). The reality of illusion. An ecological approach to cognitive film theory. Carbondale/Edwardsville: Southern Illinois University Press.

Balázs, B. (1961). Der Film. Werden und Wesen einer neuen Kunst. [*Film. Emergence and essence of a new art*] Wien: Globus Verlag.

Barg, W. C. (2019). Blockbuster Culture. Warum Jugendliche das Mainstream-Kino fasziniert. [*Blockbuster culture*. *Why young people are fascinated by mainstream cinema*] Berlin: Bertz + Fischer.

Burn, A. (2013). The kineikonic mode: Towards a multimodal approach to moving image media. In: *National Centre for Research Methods Working Papers* (03), 1–25.

Burn, A. (2016). Games, films and media literacy: frameworks for multimodal analysis. In: M. Knobel & C. Lankshear (Eds.), *Researching new literacies: design, theory, and data in sociocultural investigation* (pp. 1–38). New York. Peter Lang.

Dästner, C. (2005). Sprechen über Filmmusik. Der Überschuss von Bild und Musik. [*Talking about film music. The surfeit of image and music*.] In: Segeberg, H. & Schätzlein, F. (Eds.), *Sound. Zur Technologie und Ästhetik des Akustischen in den Medien* (pp. 83–95). [*On the technology and aesthetics of the acoustic in the media*] Marburg: Schüren (Schriftenreihe der Gesellschaft für Medienwissenschaft [GfM] 12). https://doi.org/10.25969/mediarep/14194#

Deleuze, G. (2015). Das Zeit-Bild. Kino 2 [*The time-image. Cinema 2*]. Frankfurt/Main: Suhrkamp.

Desplat, A. (2010). Snape to Malfoy Manor. In A. Desplat: Harry Potter and the Deathly Hallows—Part 1 (Original Motion Picture Soundtrack). Burbank: WaterTower Music, Track 2.

Dolar, M. (2002). Der Zuschauer, der zuviel wußte. [*The viewer who knew too much*] In: S. Žižek, M. Dolar, A. Zupančič, S. Pelko, M. Božovič & R. Salecl (Eds.), *Was Sie immer schon über Lacan wissen wollten und Hitchcock nie zu fragen wagten* (pp. 126–34). [*Everything you always wanted to know about Lacan but were afraid to ask Hitchcock*] Frankfurt/Main: Suhrkamp Verlag.

Dürr, S. (2018). Die Vermittlungsebene von Texten analysieren. [*Analyzing the mediation level of texts*] In: A. Schilcher & M. Pissarek (Eds.), *Auf dem Weg zur literarischen Kompetenz. Ein Modell literarischen Lernens auf semiotischer Grundlage* (pp. 229–259). [*Towards literary competence. A model of literary learning on a semiotic Basis*] Baltmannsweiler: Schneider Verlag Hohengehren.

Edwards, B. (1963). The Pink Panther. UK / US: United Artists.

Eisenstein, S., Pudwokin, W. I. & Alexandrow, G. W. (1998). Manifest zum Tonfilm. [*A manifesto for talking pictures*] In F.-J. Albersmeier (Ed.), *Texte zur Theorie des Films* (pp. 54–57). [*Texts on film theory*] Stuttgart: Phillip Reclam.

Elsaesser, T. & Hagener, M. (2007). Filmtheorie zur Einführung [*Introduction to film theory*]. Hamburg: Junius Verlag.

Götz, A. (2013). GFS [*Movie score sampler film*.] https://www.youtube.com/watch?v=5w2eKr4X9EA [l. A.: 12.06.2022]

Gorbman, C. (1987). Unheard Melodies: Narrative Film Music. Bloomington: Indiana University Press.

Grabbe, L. C. & Rupert-Kruse, P. (2013). Äquilibration und Synkretismus. Überlegungen zu einer interaktionistischen Theorie der Filmbildrezeption. [*Equilibration and syncretism: Considerations on an interactionist theory of film image reception*] In: L. C. Grabbe, P. Rupert-Kruse & N. M. Schmitz (Eds.), *Multimodale Bilder. Zur synkretistischen Struktur des Filmischen* (pp. 16–41). [*Multimodal images: On the syncretistic structure of the cinematic*] Darmstadt: Büchner-Verlag.

Harste, J. C., Breau, A., Leland, C., Lewison, M., Ociepka, A. & Vasquez, V. (2000). Supporting critical conversations. In: K. M. Pierce (Ed.), *Adventuring with books* (pp. 506–554). Urbana: National Council Teachers of English.

Heins, J. (2018). Qualitative Inhaltsanalyse. Mit dem Rüstzeug der QIA neue Wege gehen.[*Qualitative content analysis. Breaking new ground with the QIA toolkit*] In: J. M. Boelmann (Eds.), *Empirische Forschung in der Deutschdidaktik. Bd. 2. Erhebungs- und Auswertungsverfahren* (pp. 303–323). [*Empirical research in German didactics. Vol. 2. Survey and evaluation methods*] Baltmannsweiler: Schneider Verlag Hohengehren.

Kammerer, I. & Maiwald, K. (2021). Filmdidaktik Deutsch. Eine Einführung. [*Film didactics in German language teaching. An introduction*] Berlin: Erich Schmidt Verlag.

Kracauer, S. (1985): Theorie des Films. Die Errettung der äußeren Wirklichkeit. [*Theory of film. The redemption of physical reality.*] Frankfurt/Main: Suhrkamp Verlag.

Kress, G. & van Leeuwen, T. (2001). Multimodal discourse: The modes and media of contemporary communication. London: Routledge Falmer.

Kubrick, S. (1971): A Clockwork Orange. UK / US: Columbia-Warner Distributors & Warner Bros.

Kuhn, M. (2011). Film-Narratologie. Ein erzähltheoretisches Analysemodell. [*Film narratology. A narrative-theoretical analytical model*] Berlin / New York: Walter de Gruyter. https://doi.org/10.1515/9783110253559

Kurwinkel, T. & Schmerheim, P. (2013). Kinder- und Jugendfilmanalyse. [*Analyzing children's and youth films*] Konstanz / München: UVK Verlagsgesellschaft.

Langer, A. (2013). Transkribieren – Grundlagen und Regeln. [*Transcription—basics and rules*] In B. Friebertshäuser, A. Langer & A. Prengel (Eds.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (pp. 515–526). [*Handbook of qualitative research methods in educational science]* Weinheim / Basel: Beltz Juventa.

van Leeuwen, T. (1999). Speech, music, sound. London: Macmilan. https://doi.org/10.1007/978-1-349-27700-1

Lehr, S. S. (Ed.) (2008). Beauty, brains, and brawn: The construction of gender in children's literature. Portsmouth: Heinemann.

Maiwald, K. (2005). Wahrnehmung – Sprache – Beobachtung. Eine Deutschdidaktik bilddominierter Medienangebote [*Perception—language—observation. A didactics of German image-dominated media offers*] München: KoPäd.

Mayring, P. (2009). Qualitative Inhaltsanalyse. [*Qualitative content analysis*] In U. Flick, E. von Kardorff & I. Steinke (Eds.), *Qualitative Forschung. Ein Handbuch* (pp. 468–475). [*Qualitative research. A handbook*.] Hamburg: Rowohlt Taschenbuch Verlag. https://doi.org/10.1007/978-3-8349-9441-7_42

Mayring, P. (2022). Qualitative Inhaltsanalyse. Grundlagen und Techniken. [*Qualitative content analysis. Fundamentals and techniques*] Weinheim: Beltz Verlagsgruppe.

Medienpädagogischer Forschungsverbund (2018). KIM-Studie. Kindheit, Internet, Medien.

Basisuntersuchung zum Medienumgang 6- bis 13-jähriger [*The KIM Study. Childhood, internet, media. Basic study on media use by 6- to 13-year-olds.*] https://www.mpfs.de/studien/kim-studie/2018/ l. a. [05.06.2023].

Merlin, D. (2020). Filme hören. Auditive Dimensionen der Filmrezeption am Beispiel ausgewählter Szenen aus [Listening to films. Auditory dimensions of film reception using the example of selected scenes from] Don't Come Knocking (2005) and The Lone Ranger (2013, TV 1949). In A. Kagelmann, M. Knopp, A. Mrichel, A. Meteling & F. Münschke (Edts.), MiDU (Medien im Deutschunterricht). Multimodales Erzählen im Deutschunterricht II: Schrift – Bild – Ton, 2 [MiDU (Media in German Language Teaching). Multimodal narration in German lessons II: Writing—image—sound,] (2), 1–18. https://doi.org/ 10.18716/OJS/MIDU/2020.2.5 2

Moormann, P. (2009). Einleitung. In P. Moormann (Eds.), *Klassiker der Filmmusik* (pp. 9–18). [*Introduction. In P. Moormann: Classics of film music.*] Stuttgart: Philipp Reclam.

Oates, J. C. (1991): Dracula (Tod Browning, 1931): The Vampire's Secret. In *Southwest Review. 76* (4), 498–510.

Pauli, H. (1981). Filmmusik: Stummfilm. [*Film score: silent film*]Stuttgart: Ernst Klett.

Pissarek, M. (2018). Merkmale der Figur erkennen und interpretieren. [*Recognizing and interpreting characteristics of the character*] In A. Schilcher & M. Pissarek (Eds.), *Auf dem Weg zur literarischen Kompetenz. Ein Modell literarischen Lernens auf semiotischer Grundlage* (pp. 135–158). [*Toward literary competence. A model of literary learning on a semiotic basis*] Baltmannsweiler: Schneider Verlag Hohengehren.

Pietsch, V. (2018). Verfolgungsjagden. Zur Diskursgeschichte der Medienkonkurrenz zwischen Literatur und Film. [*Chase sequences. On the discourse history of media competition between literature and film*] Bielefeld: transcript. https://doi.org/10.1515/9783839433867

Pietsch, V. (2022). Verstehen auf Niveau? Zur Modellierung von Lernprozessen und ihrer medialen Unterwanderung. [*Understanding on a high level? On the modelling of learning processes and its medial undermining*] In H. Heimböckel & J. Pavlik (Edts.), Ästhetisches Verstehen und Nichtverstehen. Aktuelle Zugänge in Literatur- und Mediendidaktik. [*Aesthetic understanding and misunderstanding. Current apporaches in literature and media education*] Bielefeld: transcript. https://doi.org/10.1515/ 9783839460733-009

Rabenalt, P. (2014). Der Klang des Films. Dramaturgie und Geschichte des Filmtons. [*The sound of film. Dramaturgy and history of film sound*] Berlin: Alexander Verlag.

Scheubeck, T. (2023). Mit audiovisuellen Texten der Populärkultur literarische Kompetenz fördern. Das Potenzial von Werbespots im Literaturunterricht. [*Promoting literary competence with audiovisual texts of popular culture: The potential of commercials in literature education*] Münster / New York: Waxmann.

Schilcher, A. (2018). Handlungsverläufe beschreiben und interpretieren. [*Describing and interpreting plot developments*] In A. Schilcher & M. Pissarek (Eds.), *Auf dem Weg zur literarischen Kompetenz. Ein Modell literarischen Lernens auf semiotischer Grundlage* (pp. 199–228). [*Towards literary competence. A model of literary learning on a semiotic basis*] Baltmannsweiler: Schneider Verlag Hohengehren.

Schilcher, A. & Pissarek, M. (2018): Kompetenzmodell Literarisches Lernen. [*Competence model literary learning*] In: A. Schilcher & M. Pissarek (Eds.), *Auf dem Weg zur literarischen Kompetenz. Ein Modell literarischen Lernens auf semiotischer Grundlage* (pp. 324–325). [*Towards literary competence. A model of literary learning on a semiotic basis*] Baltmannsweiler: Schneider Verlag Hohengehren.

Schmitz, U. (2016): Multimodale Texttypologie. [*Multimodal text typology*] In: N. Klug & H. Stöckl (Eds.), *Handbuch Sprache im multimodalen Kontext* (pp. 327–347). [*Handbook of language in a multimodal context*] Bd. 7. Berlin / New York: De Gruyter. https://doi.org/10.1515/9783110296099-015

Schneider, E. (2006). Handbuch Filmmusik I: Musikdramaturgie im Neuen Deutschen Film. [*Handbook of film music I: Music dramaturgy in the new German cinema*] Konstanz: UVK Verlagsgesellschaft.
Seel, M. (2013). Die Künste des Kinos [*Cinema arts*]. Frankfurt/Main: S. Fischer.

Seifert, L. (2019). Filmmusik in der Grundschule. Durchführung und Evaluation eines Praxisprojekts in einer vierten Grundschulklasse. [*Film music in elementary school. Implementation and evaluation of a hands-on project in a fourth grade elementary school class*] Universität Hildesheim: Unpublished thesis.
Serafini, F. (2014). Reading the visual: An introduction to teaching multimodal literacy. New York: Teachers College Press.
Serafini, F. (2015). Multimodal literacy: From theories to practices. In *Language Arts* (92), 412–423. https://doi.org/10.58680/la201527389
Soderbergh, S. (2001). Ocean's Eleven. US: Warner Bros. Pictures.
Spadoni, R. (2007): Uncanny bodies: The coming of sound film and the origins of the horror genre. Berkeley / Los Angeles: University of California Press. https://doi.org/10.1525/9780520940703
Spinner, K. H. (2006): Literarisches Lernen. [*Literary learning*] In: *Praxis Deutsch 33* (200), 6–16.
Truffaut, F. (2009). Mr. Hitchcock, wie haben Sie das gemacht? [*Hitchcock/Truffaut*] München: Wilhelm Heyne Verlag.
Williams, J. (2004). A Window to the Past. In J. Williams: Harry Potter and the Prisoner of Azkaban (Original Motion Picture Soundtrack). Burbank / New York City: Warner Sunset, Nonesuch & Atlantic, Track 7.