The cognitive roots of multimodal symbolic forms with an analysis of multimodality in movies

Wolfgang Wildgen*

Institut für Allgemeine und Angewandte Sprachwisenschaft (IAAS), Sprach- und Literaturwissenschaften, Universität Bremen, Breman, Germany

Condillac's (1754) "Traité des sensations" is the philosophical background of modern discussions on the relationship between perception and multimodal communication. The differences between perception and communication and the transitions between them are discussed with a focus on odor and color. It becomes clear that even at this primary level, the complex interactions of different modalities are the precondition for effective and rich communication. The second part discusses Cassirer's "Philosophy of Symbolic Forms" as a relevant framework for multimodality studies. Basic aspects are first commented on with a focus on music and visual art. The interaction is even more complex and rich in the case of language; the difficulty of large symbolic forms is mainly due to semantic composition and only to a lesser degree to syntactic concatenation. The first must merge/blend different semantic spaces. It must allow for the plurality of levels of integration from the lexical level, the level of phrases and sentences, up to texts and discourse. The third part focuses on multimodality in film. It treats the representation of movement and action in (film) narratives, the visual perception and representation/communication of movement and action, and the integration of music, moving images, and language.

KEYWORDS

multimodality, symbolic forms, cognitive roots, music, visual artifacts, language, movies, semiotics

1 Sensation, knowledge, and modes of communication

The relationship between sensation and knowledge was the focus of the "Traité des Sensations" ("Treatise on Sensations"; Condillac 1754/1970). It presents the following thought experiment. A marble statue is successively endowed with sensuous experience. Condillac begins with olfaction. The statue first is what it smells; second, it distinguishes pleasant and unpleasant odors. Finally, comparison and memory select and stabilize the attention movement after the stimulus' reception. The difference is passive if motivated by a stimulus and active if motivated by memory (*cf.* Condillac, 1754/1970: 49). With the two principles of pleasure/aversion and memory/comparison, Condillac treats hearing, taste, and finally, sight, consisting of light and color. He suggests a system of ideas based on olfaction. In his earlier treatise "Essai Sur l'Origine des Connaissances Humaines" (An Essay on the Origin of Human Knowledge; 1746/1982), he tried to reduce the theory put forward by Locke (1690/1975) in his "An Essay Concerning Human Understanding" to one principle (sensation) and to explain the rise of human language and culture.

From the present perspective, the "sensualistic" position of Condillac was a sound synthesis of ideas put forward in the discussions of modernity (sixteenth to eighteenth

century) but underestimated the problem's complexity. The composition of larger structures in visual artifacts, musical composition, and text/literature asks for large-scale organizational principles, which may be called the "architecture" of human communication.

1.1 The complexity of symbolic forms beyond the senses and their integration

Despite all advances, the following significant problems could not be solved:

- How are meanings grounded in our (subconscious) sensory activities, which are adapted by evolution to selected features of our environment? The quality of this grounding (the "correct" selection, the stable transfer of relevant structural relations up to the highest levels of cognition) is crucial for the functional fitness of the "animal symbolicum," as humans are called by Cassirer (1995).
- How can individual thinking (as representation) be economical and still represent external reality?

Condillac considered perception as the first semiotic level and signs/semiotic entities necessary to stabilize and organize memory and thinking as a further one.[1] However, this scale of an unfolding of human communication is incomplete because a fundamental conflict exists between individual perception/action and social communication. The first field has the individual brain (including the sensory organs) and the individual development (maturation and essential learning in a suitable environment) as its domain. The second field concerns the adaptation to a culturally transmitted system of linguistic rules, social beliefs, knowledge, social action, and perception.

Although learning and socialization link both domains, the fundamental mechanisms differ. The organization of the brain cannot be identified with the organization of a community, and communication underlies other conditions of coding, transmission, functional adequacy, and economy than that operative in sensory organs and cortical centers of perception and motor planning. The differences become relevant in the case of olfaction and color and, thus, the contribution of sensory experience to unfolding symbolic forms.[2] Therefore, we shall consider this critical transition decisive for

---

1   One must distinguish the very fast and automatic reaction of the sense organs to the stimulus and the elaborated reactions in the sense-related cortical areas (for instance V1 in the visual cortex). In the following this elaborated stage is understood as "perception."

2   *Cf.* Wildgen (2023, 197-200) for a comparison of the notion "symbolic form" in Cassirer and the notion of "sign" (symbol) in Peirce. "Cassirer's philosophy aims at a philosophy of human culture that addresses symbolic forms such as language, myth, science (later art, ethics, and technology). The meaning of the symbolic forms is given by the internal "reference structures" of consciousness; i.e., mental entities of different provenience and character are brought together. This trend is fully developed in human cultures and concentrated in a plurality of symbolic forms. The analysis of this plurality and the interrelations between the basic types is at the heart of Cassirer's theory of culture."

evolution of human societies as a first step.[3] In the second step, we analyze complex communication and multimodal symbolic forms. Eventually, multimodal communication is studied in the interaction of visual, textual, and musical communication in movies.

1.2 The mode of communication about odors

If both the speaker and the audience are exposed to the same smell, it is enough to call the audience's attention to the scent and then perform a speech act referring to the (supposed) common perception.[4] It is only if the odor is not present and must be represented that the problem of proper categorization and characterization occurs.[5] It presupposes a stable system of meanings grounded in similar olfactory experiences. In a community where this demand does not occur frequently, communication regarding odors will be very insecure, unstable, poor, or meaningless. Thus, in the extreme case, every participant in a conversation may be perfectly able to perceive and, if necessary, distinguish a large set of odors but unable to communicate these distinctions. The required ability concerns first the stable grounding in personal perceptual experience, second efficient coordination with the perceptual patterns in those with whom one wants to communicate, and third, the invention and use of labels (lexical or periphrastic). Thus, it does not help to learn words for

---

3   The evolutionary sequence where basic capacities of perception and motor control were developed (e.g., in the Cambrian revolution) reappears in the rooting of higher mental capacities (e.g., language and other symbolic forms) in perception and motor control.

4   The notions "mode" and "modality" used in this contribution always imply communication; "modality" is an abstraction regarding modes of communication. In the present study, communication is primarily human communication, not between animals, plants, cells, or bodily substances, as in zoosemiotics. Cassirer's term "symbolic forms" is more specific insofar as he assumes that the symbolic capacity is a defining feature of humans. Since the first half of the last century, when Cassirer developed his "Philosophy of Symbolic Forms," our knowledge of the evolution of primates and humans was further advanced; therefore, hypotheses assuming the isolation of human symbolic capacity must be revised. In our context, communication is restricted to humans, and the question of generalizing our analysis beyond humans is not touched. The list of "modes" varies with the authors. Kress (2010:79) mentions: "image, writing, layout, gesture, speech, moving image, soundtrack." He refers to "meaning-making" as a common feature. The crux with "meaning making" as the central criterion is that the traditional notion of "meaning" in linguistics does not match parallel features called signification, significance, or functional effect in painting, architecture, music, and all modalities different from language. In this essay, we distinguish primary modes (of communication) linked to human senses: touch, odor, sight, and hearing. Consequently, language is considered a secondary mode, mixing different sense-related modes or being dissociated from these modes. Speech and writing would be modes dominated by the linguistic mode.

5   Sections 1.2 and 1.3 use materials written by the author to prepare the monograph Wildgen (2023). Passages in this contribution correspond literally or in content to passages in sections 2.2 and 2.3, p. 17–22 of Wildgen (2023). This general advice must be sufficient, as explicitly marking all parallels and differences would have impeded the reading.

odors. Even if these labels are individually grounded in olfactory experience, olfactory communication remains vague if no social coordination has been achieved. The grounding of odors is difficult because the chemical structure of odors is very complex, and behavioral reactions are very context-dependent. Nothing corresponds to the Munsell classification of colors in the domain of odors (*cf.* Dubois, 2021: 203f).

Condillac (1754/1962) mentioned the problem of organization of the field of odors and the fact that it depends on the conditions and contexts of memorization.[4] Thus, if one experiences a series of odors, e.g., of flowers, coffee, and a nearby factory, a network that links and partially mixes these impressions and their evaluative reactions is created in memory. As a result, the neighborhood of odors in time and space, the amount of attention paid to them, and the pleasure or displeasure associated with them change from one individual to the next.

To summarize, the inventory of linguistic devices to account for 'odors' as a sensory experience, when extended beyond the search for 'basic terms', reveals that olfaction is not a 'mute' sense. The lack of lexical items in olfaction depends on sociocultural and linguistic constraints of 'talking about odors' rather than physiological constraints. A large diversity of morpho-syntactic, syntactic, and discursive devices is at work, differing from the (predominant lexical) categories relevant to vision (Dubois, 2021: 229).

1.3 The mode of communication referring to colors

Condillac 1754/1970 argues that the perception of colors alone does not constitute colored places, situations, motions, or objects. Instead, our visual brain allows us to see two or three colors and journey from one color to another. Eventually, the perception of colors contributes to an idea of extension, a space of colors.[6]

The astonishing differentiation of human color vision does not automatically entail a rich system of "ideas" of colored surfaces in our minds. Instead, attention, memory, the color space (surface) organization, and paths of visual attention are necessary to produce a selected set of "color ideas". Higher levels are only achieved if other types of sensibility are integrated and form an organization linked to objects, events, and actions. As the divergence of color terminologies in the world's languages shows, the need for a basic (stable) set of color terms varies between languages/ethnic groups, even on a similar level of cultural development. Therefore, communication about colors is not a universal and central concern of linguistic communities. However, this restriction does not preclude that the cognitive relevance of color perception is very high in individual perception, i.e., perception and communication do not share the same functional

pattern and do not respond to identical (or similar) needs. Instead, they follow their specific principles and laws.

In anthropological linguistics, Berlin and Kay (1969) have examined the elementary color vocabularies of different language cultures and the color values assigned to color words. A comparison with the physiological data resulted in a differentiation hierarchy. Color physiology can foresee the options if the degree of differentiation of color terminology increases in one language or in the case of a transition from one language to another.[7] One can infer that, to a certain extent, culture-invariant color vision determines the basic color lexicon. This result speaks for a weak cognitive determinism.

Although the universality of color terms or their sequence of replenishment remains controversial, it became clear that putting the perception of colors (in context) in words or utterances is a complex activity that is, in many cases, successful. Therefore, skepticism regarding the linguistic realization of primary color perception is not warranted. Instead, color perception and the perception of spatial contours, lines, and other visual features can achieve the status of symbolic forms in a semiosis that uses more general resources.

2 Multimodal integration and symbolic forms

In a given linguistic community, some sub-communities, i.e., painters or perfumes professionals, develop specific teaching competencies and techniques or even argue about odors/perfumes or colors. This practice modifies the communication demands. Our reflections point to three significant problems or problematic transitions between "sensibility" (perception) and "sense" (meaning in a symbolic form, e.g., language):

a The transition from single sensibilities, e.g., olfaction or color sensibility, to an integrated perception is crucial.
b The transition between presentation and representation must consider the effect of attention, memory, and spontaneous imagination.
c The transition between perception, governed by principles of human neural architecture, and the dynamics of human communication, based on social interaction.

The most dramatic transition and the evolutionarily most recent one is that of (b) to (c). At level (c), specific human capacities are concentrated in the plurality of "symbolic forms" discussed by Ernst Cassirer. The problem of multimodality has been noted and discussed in philosophy at least since the 18th century; see our remark on Condillac in section 1. In the twentieth century, the philosopher and historian of science Ernst Cassirer developed a philosophy of symbolic forms and human culture. His notion of the symbolic forms is a good background for theories of multimodality insofar as the relative autarky of different modes and their interaction is put to the fore. Although the symbolic forms of language and theoretical/ mathematical form-giving in the sciences are considered exemplary,

---

6   In Petitot (2017: 266) the neural integration of retinal opponent cell operations and positional information, mainly spatial frontiers is discussed in detail. "So from the lowest levels of the V 1 and the V 2 areas, there is a functional entanglement between the spatiality of perceived scenes and the colours of objects making them up. The brain must reconstruct by inference from the colour the objective reflectance of the surfaces perceived and this independently of a host of extremely variable factors such as the illumination of sources, indirect irradiation, angles coincidence, and reflection."

---

7   The color project was continued by Kay and others. *Cf.* for a later stage: Kay and Maffi (2013).

the other symbolic forms, e.g., art, myth, technology, ethics, and further ones, exist in parallel and may be older and thus the evolutionary sources of language and science. Therefore, we shall discuss the role of language on behalf of the more basic forms like odor- or color communication with specific reference to Cassirer's notion of "symbolic form."

2.1 What are symbolic forms?

Cassirer introduced "symbolic form" in his "Philosophy of Symbolic Forms," published in three volumes between 1923 and 1929. It contains two constituents: "symbolic" and "form." The second term includes the notions of morphè (Greek: shape/form) and its dynamic counterpart morphogenesis. The first term, symbolic, refers to symbol, a polysemic notion used in philosophy and aesthetics. A discussion of the range of meanings of this notion would ask for an independent treatise. However, an appropriate starting point is the notion of symbol in the theory of signs proposed by Charles Sanders Peirce in the second half of the nineteenth century. *Cf.* for comparing the contributions of Peirce and Cassirer, Wildgen (2023, Chap. 7).

Peirce considers three fundamental aspects of the sign, the central notion of semiotics: icon, index, and symbol. In this constellation, the notion of the symbol has to be delimited concerning the two neighboring concepts of the index (existential reference) and icon (reference via similarity, conceptual neighborhood).[8]

This definition needs clarification. The "dynamic object" is, in Peirce's terminology, the real-world object, the ultimate intention of the sign in its real-world usage. Peirce's statement tells us that the symbol only indirectly relates to the real-world object ("in the sense it is interpreted"). The interpretation depends on dispositions, habits, or conventions (see fn. 5). These determinations introduce a moment of arbitrariness; they depend on chance or, eventually, on many minimal causes beyond rational control. Regarding dispositions and habits mentioned by Peirce, one can assume rules of behavior that were acquired but gained law-like significance. In the case of conventions, these forces include cooperative effects in a community and conformity in social behavior.

2.2 Complex symbolic forms: the example of music and visual art

Immediate perceptual processes, such as those observable in olfaction and color vision, are still near to natural (bodily) morphogenesis. In contrast, colored surfaces and objects or artifacts (art) are symbols in the sense of the definition given by Peirce. In the case of odors, smell, and taste, very specific, institutionalized, and professional situations or contexts can lead to artificial norms and

devised terminologies that produce a symbolic level on which such perceptions may be efficiently communicated.

The symbolic forms of music and visual artifacts have a dominant founding in specific physical conditions: auditory perception and the motoric capacities to produce music in the first case and visual perception and motoric capabilities for creating visual artifacts in the second case. Language, an evolutionary late-comer, has a broad field of interacting capacities and needs many sources. The symbolic forms of music and visual forms can use the complexities of language to elaborate their repertoire and symbolic richness culturally. The multimodality of visual, acoustic, and linguistic communication is a major concern in social semiotics based on the linguistics of Halliday and further developed by van Leeuwen, Kress, Bateman, Wildfeuer, Hiippala, and others.[9]

The list of symbolic forms or communication modalities may be subdivided, although the evolutionary continuity implies an underlying continuum. As we showed in section one, odor/smell and basic color distinctions are at the threshold of symbolic forms. Visual artifacts and musical performances are still firmly rooted in perceptual patterns and motoric routines, i.e., bodily controlled (embodied). In contrast, languages are rooted in different perceptual domains, emerge or co-evolve from musical behavior (*cf.* Wildgen, 2018: 62–78), and refer to visually rooted (virtual) spaces. These levels range from the perceptually dominated odor/smell, still bodily rooted visual and musical forms to the more abstract and, to a large degree, culturally transmitted and sophisticated linguistic forms. Beyond, we find secondary symbolic forms heavily dependent on the three layers mentioned above but with new functions and more specific effects, for instance, myth and religion (*cf.* Wildgen, 2021) and the symbolic forms, technology, and ethics that Cassirer has added to his list of symbolic forms (*cf.* Sandkühler et al., 2003: 34f, 42f, and chapter 9). In the analysis of multimodality, it seems primary to consider the visual, musical, and linguistic symbolic forms. Multimodal artifacts or performances may combine or integrate two or three of these forms:

- The couple: visual and musical forms. Simple analogies concern the visual ornament in paintings or architecture and the ornamental enrichment of a melody. *Cf.* for examples and analyses Wildgen (2018: 97-102). Some painters like Paul Klee (1879–1940) have reflected the analogies between music and graphic or color design (*cf. ibidem*: 102–106).
- The couple: visual and linguistic forms. The close relationship is not only demonstrated by the evolution of writing based on pictorial representations but also on parallel vocations, i.e., painters who are poets and poets who are painters. Principles of symmetry and spatial ordering are valid for visual artifacts and poetry (*cf.* Wildgen, 2013: chapter 10). Stories and literary fiction can be realized via language (spoken or written), illustrations or sequences of pictures (*cf.* comic strips), and movies.
- The couple: musical and linguistic forms. The simplest integrated form is given with songs combining melody and lyrics. Poetry, with its rhymes and alliterations, the "concert" of vowels and consonants, and rhythmic features rival music or interfere with it.

---

8   In a letter to Lady Welby on October 12, 1904, Peirce summarizes his position (developed after 1867, as he tells her; *cf.* Wiener (1958: 391). "I define a Symbol as a sign which is determined by its dynamic object only in the sense that it will be so interpreted. It thus depends either upon a convention, a habit, or a natural disposition of its interpretant (that of which the interpretant is a determination)." (*ibid.* 291f).

---

9   For an overview of web design applications, comics, film, audio-visual materials, and video games, *cf.* Bateman et al. (2017) and an introduction to social semiotics Kress and van Leeuwen (1996).

- The triple: music, visual art, and language is typical for performances of musicians on stage, opera, and musicals and, since the 19th century, for movies. This topic will be the main concern in the section 3.

2.3 Language depends on the self-organization of perceptual capacities and their multimodal integration

Self-organization is a principle formulated in the cybernetics framework (Ashby, 1947) and involves searching for a stable state in a deterministic system. As already programmatically expressed by Wiener (1951), it is extrapolated from physical to biological, eventually symbolic systems. The purely syntactic problem of chaining elements of an existent vocabulary does not require a specific endowment and evolutionary processes enabling it; it can exploit much older sequencing techniques in motion and action. The real problem is semantic compositionality because the composition or blending of spaces with different topologies and the account of verb dynamics is crucial for sentential units. This tremendous problem must be resolved to allow stable and reliable communication via phrases and sentences. To arrive at a conventionalized system of linguistic behavior, early humans had to consider two major factors:

a The cognitive demands for a stable solution of semantic compositionality,
b The communicative and social demands for a compositional level of referentiality.

The solution to this problem is the gain of the evolutionary game called human language. Human utterances are, however, not restricted to isolated sentences. On the contrary, natural units are sequences of sentences, turns in conversation, adjacent pairs as in question–answer, and narratives or arguments. Therefore, human evolution has created the human language for its effective use in social communication, not for correctly using sentences or words.[10]

3 Multimodality in movies

Movies are an intriguingly complex example of multimodality, historically only comparable to theater and opera. In the following, we recapitulate major results obtained in analyzing movies in Wildgen

(2015, 2017, in English, 2013: chapter 7, and, 2018: 107-112; in German).

3.1 Space, movement, and narrative in movies

Movies are, on the one hand, moving images and visual communications in time. On the other hand, a story is being told. The balance between visual attraction and narratively motivated action differs in film genres. In action films, the focus is on the dramatic action focused on just a few actors, but it must be "woven" into a narrative texture. The public perceives many action films as part of a film series (see the series of James Bond movies). The narrative thread must, therefore, point beyond the respective film, provide structural analogies to previous films, and possibly prepare the next film.

3.2 The visual mode of movement in space

The location of the plot is the anchor for what is happening in the movie and makes it appear believable. In addition, characters and actions only become understandable and effective as constructs in the context of these places. In a broader sense, locations also include the costumes and location-specific behavior of the characters. In this respect, the film's basic structure is already established with the exact construction of the locations and the directing of the courses of events and actions at these locations. Places of transition also play a major role, such as hotel lobbies, elevators, train stations, airports, and crowded squares (cf. the analysis of James Bond movies in Wildgen (2015, 2017).

The frame, the viewing window, plays a decisive role. A film in the Academy format (square) emphasizes the center more, thus increasing the illusion of depth. The broadband format emphasizes the horizontal, giving the landscape and storyline greater prominence. Actions and movements across the camera window can be followed longer without changing the setting. For example, suppose the film is set in architectural interiors. In that case, elements of architecture: doors, staircases, windows, narrow corridors, room dividers, and even furniture can create specific frames within the format and thus shape the space structure. People can be assigned to individual room segments. These spatial divisions can be repeated when moving through a suite of rooms. The person's (and the camera's) line of sight can be downwards (from a balcony, an upper floor window into the yard, onto the street) or up into a stairwell or, particularly extreme, into a rock face when climbing (correspondingly down into the dizzying abyss). In connection with the division of space, people and their actions acquire specific meanings. Structuring the space (particularly through the dividing lines and thresholds) is meaningful since it creates spatially separated binding structures of thematically related subfields (cf. Saint-Martin, 1990: 208ff). In the film, the spatial structures are transformed by the movement of the people and the camera. The film can even be viewed as a medium of spatial transformation. The moving person can be focused in the foreground; the surrounding space flows past the person. This feature is particularly evident in older Hollywood films and in some films of the New Wave, where the actors are filmed at the wheel of a car over long takes. However, the movement can also be caused by the camera

---

10 The correctness criterion, traditional for school grammar, is linked to social conformity (obedience to rules) and sharply delimited social identities. Chomsky's notions of grammaticality and the construction of "competence in a native speaker" that enables judgments of grammaticality is an abstraction that neglects the content of utterances in favor of formal features of concatenation (syntax). The intuitive notion of acceptability used by Chomsky's teacher, Zellig Harris, was much easier to measure empirically and was nearer to the traditional notion of correctness. Indirectly, one could argue that Chomsky's notion of competence is akin to traditional (prescriptive) school grammar. Cf. for current research on the notion of competence Vulchanov et al. (2022).

moving or the setting being varied from a long shot to a close-up. The cameraman plays a significant part in constructing meaning. The sequence of scenes and actions in different spatial segments is performed in the montage (in the editing room). Camera settings and the construction of the sequence of scenes in the editing room are thus the main organizational level of cinematic meanings. We can thus distinguish three sub-levels of the meaning construction of characters in the film:

1 The construction of meaning in the scene in front of the camera (prepared in the script, planned and controlled by the director, and specified by the actors).
2 The construction of meaning through the camera bears on the choice of setting, control of the lighting effects, and the camera's movement in space. In most cases, a multiple of the required film material is recorded, i.e., the camera creates a potential narrative space from which radical selections are made. Finally, complementary to the captured image is the off or hors-champ, which can be connotative.
3 The assembly of ready-made parts is either privative, i.e., large parts of the film material are discarded. The film director in the editing room corresponds to the sculptor who shapes the product that only exists in the imagination from a block of marble. Alternatively, the narrative order is created in the montage. In contemporary films, the scenes are reworked using computer technology and supplemented with special effects. In other cases, the main components of the entire film are generated electronically and supplemented by scenes recorded by a camera (the movements of real actors then serve as material for the animation of artificial characters, which are realized on the computer).

These three levels of meaning are essentially optical and visual. The textual-linguistic and the musical-acoustic dimensions can be added to them. Various versions can also be made with other texts in other languages or without integrated music, e.g., performed live by an orchestra. This autonomy of the different modalities makes the separability of the three basic levels, image — text – music, clear. As shown above, the visual level of organization is broken down into the organization in front of the camera (director's and actor's performance) – camera and lighting performance — editing, assembly, and special effects. The film must integrate these three levels and their sub-levels, i.e., put them together without too much redundancy and avoid disaccord or inconsistency. The integration occurs in special zones of the respective organizational level, so these remain relatively autonomous. The montage and the text organization must fit together. The modification by montage can change the narrative content and, thus, the textual level. The camera's focus must also be in harmony with a person's weight or role in the text. Suppose the main character is not emphasized visually by size or sharpness or in the movement. In that case, certain narrative threads are not adequately realized in which she/he is the center of attention. The integration of music must be coordinated with editing and montage. However, it is also tied to the narrative structure insofar as complication and climax passages of the narration correlate with the music. The dominant dimension is the visual construction, which the actors, camera, editing, and montage carry out. In addition to the continuous (imperceptible) cuts in Hollywood films, one can consider Eisenstein's formalistic montage, Orson Wells' non-causal montage, and Godard's montage of the gap (*cf.* Agotai, 2007: 98).

3.3 Film music or the integration of music, (moving) images, and language

The music in the film can form a background without much effect and be devoid of any informational content.[11] The history of film, music, and literature shows:

- The film can be received without language (*cf.* the experience of silent movies).
- The music can also be satisfactory without speech, song, or accompanying text (*cf.* instrumental and electronic music).
- The text of classical literature can do without illustration by pictures (or even films) and musical accompaniment.

The central question will be how the visual medium on the one side and the narrative and descriptive dimension of the film on the other relate to the music's temporal-rhythmic and harmonic-melodic structure. The relationship can be a juxtaposition (an additive combination) or a selective interlocking. The rhythm of the scenes or cuts in the film and the phrases or melodic lines of the music can motivate a creative composition with emerging new qualities. The interaction can be parallel or contrary (contrapuntal). Within the three-fold relationship, music–film–language, two-way relationships can also appear. Music and language can be represented by songs, which are then embedded in the film or even form the main theme. Filmed dialogues integrate image and language and can be embedded in a musical context. The film can also have a musical or an opera as the subject and show a combination of music, language (e.g., in songs), and visually presented actions. Finally, the film can refer to other films or film-making, and the music of another film can be quoted.[12] Different combinations or blends can be conceived:

1 *The juxtaposition of music and film*: in the silent film era, the effect of the darkened rooms, the ghostly light displays, and the projectors' noise could be masked and mitigated by the music playing simultaneously (see Adorno and Eisler, 1944/1977: 116). The music used was mostly classical-romantic piano and salon music (*cf.* Kreuzer, 2001: 26). Both in the visual and in the musical area, a conventionalization of the means of expression leads to the cliché, i.e., the means are known from other contexts and uses and thus lose their current

---

11 A chapter in Wildgen (2018: 107-112; in German) treats the role and development of music in movies. The monograph discusses general and detailed aspects of the interaction of music with language and art.

12 Piel et al. (2008: 65f) refer to the film: First Name Carmen (original title: Prénom Carmen) by Jean-Luc Godart (in 1988). The protagonists pretend to be making a movie (but are planning a bank robbery). In the film, there is an orchestra whose violist is, in turn, a character in the film. Beethoven's string quartet, which is being rehearsed by the orchestra, also provides the film's structure. Even more often, such a cross-reference between music and film is witnessed in films about musicians.

expressiveness. From this perspective, Adorno and Eisler (1994/2007) criticize the film industry (especially in Hollywood before 1930). The reuse of motifs from the classical music of the 19th century does not correspond to the historical context of the film viewer, as it reflects intellectual and aesthetic movements in the 19th century. The firmly established classical musical forms used in different cinematic and narrative contexts are redundant or meaningless in these contexts.[13]

2 *The partial integration of film and music*: it resulted from the desire to combine scenic/visual and musical expression. From 1909, technical and commercially viable proposals for solving the problem came into use. Types of scenes in the film were associated with musical examples with corresponding dynamics or an appropriate pace. In the successful era of Hollywood films (1908–1927), "cue sheets" appeared, i.e., a script that assigned music scores to the film's scenes. The composers began composing their music adapted to the individual film.[14] The elements of the musical tradition that were processed for the film mostly came from Richard Wagner, Richard Strauss, Giacomo Puccini, Giuseppe Verdi, Gustav Mahler, Claude Debussy, Maurice Ravel, and Alexander Scriabin. The technique of integrating music and cinematic events first resorted to the technique of leitmotifs that Richard Wagner had developed for the opera.[15] Recurring motifs or melodies are assigned to important characters or locations, thus reinforcing the movie's cinematic and textual coherence.[16] Max Steiner introduced the accompanying music to the dialogues. Specific solutions had to be found for the technical problem of an exact adjustment of the film and musical sequences, e.g., the use of stop signals for the orchestra and the conductor of the film music or even the writing of the film music according to a stopwatch. When both the film and the soundtrack could be cut and put together in montages, and especially since the components of the film were ordered on the computer or since digital production, there was no technical obstacle to the temporal coordination of film and music (through so-called temp tracks).

3 *The complex integration of music and film*: instead of supplementing the sequence of images with a sequence of musical sequences, the music can also be organized in opposition to the image, as in Hitchcock's film Rebecca (USA 1941), for which Franz Waxman wrote the music. The character, e.g., Rebecca, may be absent, yet the musical motif represents her. The leitmotif linked to a person or a location can be replaced by a theme that runs through the entire film and gives it unity; compare the film: "Play Me the Song of Death" (Italy/ USA 1968, director: Sergio Leone). The film music was composed by Ennio Morricone (1928–2020). The title melody on the harmonica is embedded in the sound of a symphony orchestra with a choir. Morricone thus created a new standard for film music that was valid until the digital film era.

4 *Specific musical colors and moods*: with the focus on psychological aspects in film, moods were increasingly reflected through music; even a "mood technique" was developed. In the context of "Film Noir," dissonances played a role in controlling emotions.[17] Miklós Rózsa (1907–1995) portrayed the psychological violence of criminal characters with the help of dissonant music, for example, in "Double Indemnity" (USA, 1944; director: Billy Wilder). The film music thus became more and more independent of the visually presented events. For example, the music for the film "Star Wars" (USA 1977; directed by George Lucas, music by John Williams) was largely realized as orchestral music. It claimed to present the narrative pattern parallel to but independently of the film through synthetically modulated noises. An economically desirable consequence was that the music could be marketed independently of the film.

5 *Musical innovation in film music*. Hitchcock worked closely with his composer Bernard Herrmann (1911–1975) during the production process. The subject of vertigo is represented both in the filmed action and in the music (*cf.* Kalinek, 2007). The central motif consists of arpeggios played in opposite directions by the discant and bass voices without creating a stable melody. The beginning and end of the motif are characterized by sevenths and seconds, i.e., by dissonant intervals. This pattern is later rhythmically doubled, i.e., accelerated and marked by great changes in dynamics from ff (forte forte) and pp. (piano piano). In addition, the tempo moves from accelerando to ritardando. The musical swirl's intensity is increased by shifting the focus from the first to the second beat. In one scene, the swirl (or vortex) is also rendered visually by a rotating-colored spiral, i.e., the musical structure receives a geometrical and visual equivalent.

Film music is historically dependent on the overall development of music despite its autonomy in the phase of silent movies. Under the influence of digital technologies and the emergence of easily accessible libraries of musical motifs or passages on the internet, this has become a general characteristic of modern music culture (*cf.* Vernalis, 2013). On the one hand, recycling classical-romantic music traditions

---

13 The use of musical clichés is the consequence of industrialization in film production and the commercial pressure on film composers, who had to forego the risks of artistic innovation for the sake of box office revenue. Film music, therefore, lagged decades behind the modern development of music.

14 Important composers and arrangers in Hollywood were: Max Steiner (1888–1971), Erich Wolfgang Korngold (1897–1957), and Alfred Newman (1900–1970).

15 The leitmotifs are also taken from other films. Max Steiner gave his orchestrator brief instructions as to which motifs from which films he should use (*cf.* Wegele, 2010: 20). In contrast to the early phase of the film, Hans Zimmer writes musical suites for the planned film even before shooting begins. They are then adapted to the film at the end using modern methods or the film is adapted to the music.

16 See Weill (1946/1990: 135). Film music was shaped by this successful style of the early days well into the 20th century. However, electronic music gradually dissolved this basic pattern (see the work of Klaus Zimmer, born in 1957).

17 Ennio Morricone mentions in an interview that filmgoers will likely accept dissonant music in brutal and shrill horror films. His attempts to use modern music as film music usually failed due to the directors' lack of understanding. They pointed out to him that films tend to be produced for ordinary people.

reinforced the public consciousness of classical music. On the other hand, a compilation of the most diverse musical styles became possible. Art music is mixed with pop music, and synthetic auditory sounds are combined with speech sounds.[18]

The fact that the semiotic structures of the music, like those of the (moving) image, are self-sufficient makes their connection, their semiotic "blend," particularly attractive because the transitions and the mutual complementarity allow the emergence of new meanings in the interaction of the two symbolic forms. The new media of television, video, and the Internet also benefit from this trend. However, a longer historical phase was necessary before the potential of an effective combination of film (image) and music (sound) finally led to satisfactory results and innovative developments.

4 Conclusion: multimodality and semiotics

The rivalry of arts, mainly poetry, painting, and music, has been a topic since antiquity. In modernity, Gotthold Lessing (1729–1781) focused on the rivalry between poetry and painting in his essay Laoccon. An Essay on the Limits of Painting and Poetry (1766). The underlying question of the contribution of our senses: touch, odor, sight, and hearing to human understanding had already been thoroughly treated by Etienne de Condillac (1714–1780) in two books (*cf.* section 1). Both founding fathers of semiotics (sémiologie in Saussure's terms), Ferdinand de Saussure and Charles Sanders Peirce, tried to generalize their original field of study (before and after 1900). Saussure looked beyond language, and Peirce looked beyond philosophical logic. In this move, other types of semiosis came to the fore. However, they remained secondary to language (Saussure) or in comparison to logic (as a formal language) and science (Peirce). For the founders of semiotics, the question of multimodality remained secondary. It was the philosopher Ernst Cassirer (1874–1945) who addressed the interaction and conflict of different forms of semiosis (called "symbolic forms") in his trilogy on language (vol. 1), myth (vol. 2), and science (vol. 3) in the twenties of the last century. Later, he enlarged the field to cover art, technology, and others. A major concern of this contribution is the suggestion that the modern discipline of multimodality studies may find a proper framework in the philosophical traditions from Condillac to Cassirer.

In the main sections, the interaction, rivalry, and conflict between language (text), art (painting), and music were the main topics. At the heart of it lies the question of "meaning" or "signification" in the three modalities.[19] Major questions are:

- How can "meaning" and its semiotic realization be compared across modalities? Is there a common denominator?

- How is the apparent complexity of human products in these fields organized? Are the architectures of rich "meaning compounds" comparable? How far is an integration or blend of the different modalities possible?

The answers in traditional structuralism since Bloomfield, Hjelmslev, Chomsky, and others either avoided the question of (referential) meaning because it seemed to be ontologically freighted or relegated it to some intuitive logic (Hjelmslev, 1935) or formal ontologies in the case of intensional or possible world logics. Such strategies are of no avail in the treatment of visual art or music and only a meager substitute in the case of language. The problem of "meaning" remains the heart piece of multimodality semiotics.

The surface forms (in morphology and syntax) may be reduced to simple base forms and rules of concatenation and transformation (*cf.* Chomsky, 1957). Hockett (1954) had already distinguished two techniques of linguistic analysis: Item and Arrangement (I.A.) and Item and Process (I.P.). Both can be exemplified for languages and can easily be applied to musical and visual forms, for instance, temporal sequences of musical phrases and planar or spatial organization of elements in paintings and architecture. The different modalities differ mainly in the dimensionality of the organization in space and time. The semantics of signs or sign complexes cannot be reduced to such simple principles. They establish not only a link to the complexity of nature (the world surrounding humans) but also to cognitive/mental spaces, emotional reactions (feelings), and functional practices giving sense to symbolic behavior. In these respects, the modalities are very different. Although human language has, in many respects, a denotative function embedded in emotional/cognitive and practical contexts, music can only, under specific conditions, have denotative functions, for instance, in program music. Visual artifacts (art, architecture, film) may be illustrative and thus come near to the denotative function of language. However, as demonstrated in abstract paintings or instrumental music, they are not fundamentally denotative and can be ripped off this function.

The real challenge of multimodality lies in the difficulty of blending, integrating, or comparing the meaning effects of different modalities. The spaces of meaning have different organization and content; one cannot just arrange these contents in the same manner as the elements in a linear sequence of phonemes or morphemes.[20] A further difficulty consists of the dependence of multimodal semiosis from contexts that are either shared in a culture or fixed in use situations. This aspect was highlighted in social semiotics referring to multimodality by Bateman et al. (2017).

Author contributions

WW: Writing – original draft.

---

18 A special genre of music has emerged in the context of computer games. The music takes on new functions. The American musician Garry Schyman, who composed the soundtracks for the Bioshock series, writes: "Our compositions accentuate and expand the story. The moods we create are part of the gaming experience." Graff (2017: 9, col. 4).

19 The topic of "meaning" in art, music, myth and language has been treated in several publications by the author, recently in Wildgen (2023).

---

20 *Cf.* Brandt (2004). The blending of semantic domains is only an extended device using logical operations of selective union regarding sets of features. A more realistic model should respect the topology of different domains and check the coherence of mappings between local spaces. The proper background is given by dynamic systems theory and not logics.

Funding

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AR declared a shared affiliation with the author WW to the handling editor at the time of review.

Publisher's note

References

Adorno, T. W., and Eisler, H. (1944/1977). "Komposition für den Film" in *Hanns Eisler*. Gesammelte Werke (edited by Stephanie Eisler and Manfred Grabs) (Berlin: Hernschel).

Adorno, T. W., and Eisler, H. (1994/2007). *Prejudices and Bad Habits*, reprinted in Dickinson, K. (ed.) (2007) Movie Music. *The Film Reader*. London: Routledge, 25–35.

Agotai, D. (2007). *Architekturen in Zelluloid. Der filmische Blick auf den Raum*. Bielefeld: Transcript.

Ashby, W. R. (1947). Principles of the self-organizing dynamic system. *J. Gen. Psychol.* 37, 125–128. doi: 10.1080/00221309.1947.9918144

Bateman, J. A., Wildfeuer, J., and Hiippala, T. (2017). *Multimodality foundations, research, and analysis: A problem-oriented introduction*. Berlin: De Gruyter Mouton.

Berlin, B., and Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: Berkeley U.P.

Brandt, P. A. (2004). *Spaces, domains, and meanings*. Bern: Peter Lang.

Cassirer, E. (1995). *Philosophie der symbolischen Formen. 3 vol. [first edition 1923/1925/ 1929]*. Darmstadt: Wissenschaftliche Buchgesellschaft [English new translation: The philosophy of symbolic forms. London: Routledge, 2020].

Chomsky, N. (1957). *Syntactic structures*. Den Haag: Mouton.

Condillac, E. (1746/1982). *Essai Sur L'Origine des Connaissances Humaines: Ouvrage où l'on réduit à un seul Principe tout ce qui concerne l'entendement humain*. Paris: Pierre Mortier Translation by Franklin Philip, essay on the origin of human knowledge, in: Philosophical writings of Etienne Bonnot, Abbé de Condillac, Volume II, Hillsdale NJ: Lawrence Erlbaum, 1982.

Condillac, E. (1754/1970). *Traité des sensations*. London/Paris: Bure. Slatkine reprints https://gallica.bnf.fr/ark:/12148/bpt6k96333766/f12.item.texteImage, 1970.

Dubois, D. (2021). *Sensory experiences: Exploring meaning and the senses*. Amsterdam: Benjamins.

Graff, B. *Pokémon Symphony. Der Sound von Computerspielen wird so sorgfältig komponiert wie Filmmusik*. Mit einem Unterschied", Süddeutsche Zeitung, 98 (28th of April, 2017), part 2: 9, rows 1–5. Munich: Süddeutscher Verlag, 2017.

Hjelmslev, L. (1935). *La catégorie des cas. Étude de grammaire générale. New* Edn. München: Fink.

Hockett, C. (1954). Two models of grammatical description. *Word* 10, 210–234. doi: 10.1080/00437956.1954.11659524

Kalinek, K. (2007). "The language of music" in *A brief analysis of Vertigo*. ed. K. Dickinson (London: Routledge), 15–23.

Kay, P., and Maffi, L. (2013). "Number of basic colour categories" in *The world atlas of language structures*. eds. M. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology). Available at: https://wals.info/chapter/133

Kress, G. (2010). *Multimodality: a social Semiotic approach to contemporary communication*. London: Routledge.

Kress, G. R., and van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. New York: Routledge.

Kreuzer, A. C. (2001). *Filmmusik. Geschichte und Analyse*. Frankfurt/Main: Peter Lang.

Locke, J. (1690/1975). *An essay concerning human understanding*: Oxford: Clarendon Press, First publication, Eliz Holt, London 1690; cf. Available at: https://www.gutenberg.org/files/10615/10615-h/10615-h.htm.

Petitot, J. (2017). *Elements of Neurogeometry*. Functional Architectures of Vision: Springer, Cham.

Piel, V., Holtsträter, K., and Huck, O. (Eds.) (2008). *Filmmusik. Beiträge zu ihrer Theorie und Vermittlung*. Hildesheim: Olms.

Saint-Martin, F. (1990). *Semiotics of visual language*. Indiana: Indiana University Press.

Sandkühler, H. J., Pätzold, D., Freudenberger, S., van Heusden, B., Plümacher, M., and Wildgen, W. (2003). *Kultur und Symbol*. Ein Handbuch zur Philosophie Ernst Cassirers, Stuttgart: Metzler.

Vernalis, C. (2013). *YouTube, music video, and the new digital cinema*. New York: Oxford U.P.

Vulchanov, V., Sorace, A., Suarez-Gomez, C., Guijarro-Fuentes, P., and Vulchanova, M. (2022). The notion of the native speaker put to the test: recent research advances. *Front. Psychol.* 13:875740. doi: 10.3389/fpsyg.2022.875740

Wegele, P. (2010). Max Steiner und die Filmmusik des Golden Age in Hollywood: Eine kurze Betrachtung der wichtigsten stilistischen Merkmale anhand der Musik Steiners zum Film. *Kieler Beiträge zur Filmmusikforschung* 6, 8–36. doi: 10.59056/kbzf.2010.6.p8-36

Weill, K. (1946/1990). ""Music in the movies", Harper's bazaar" in *Musik und Theater. Gesammelte Schriften. Mit einer Auswahl von Gesprächen und Interviews* (New York: Hearst).

Wiener, N. (1951). *Cybernetics: or control and communication in the animal and the machine*. Paris: Hermann & Cie.

Wiener, P. P. (Ed.) (1958). *Charles Sanders Peirce*. Selected Writings, New York: Dover Publications.

Wildgen, W. (2013). *Visuelle Semiotik: Die Entfaltung des Sichtbaren: Vom Höhlenbild bis zur modernen Stadt*. Bielefeld: transcript-Verlag.

Wildgen, Wolfgang, Catastrophe theory and Semiophysics: with an application to movie physics, *Language and semiotic studies 1*: 61–88 (Soochow University Press, China, (2015).

Wildgen, W. (2017). "'Movie Physics' or dynamic patterns as the skeleton of movies" in *Film text analysis (Routledge advances in film studies)*. eds. J. Wildfeuer and J. A. Bateman (London: Routledge), 66–93.

Wildgen, W. (2018). *Musiksemiotik: musikalische Zeichen, Kognition und Sprache*. Würzburg: Königshausen & Neumann.

Wildgen, W. (2021). *Mythos und Religion: semiotik der Transzendenz*. Würzburg: Königshausen & Neumann.

Wildgen, W. (2023). *Morphogenesis of symbolic forms: meaning in music, art, religion, and language*. Cham (C.H.): Springer Nature (Series: Lecture Notes in Morphogenesis),.