Martin Oja*

Conflicting modalities in feature film: from contrapuntal editing to internal diegetic sound

https://doi.org/10.1515/sem-2023-0068
Received May 17, 2023; accepted May 28, 2024; published online June 12, 2024

Abstract: This article approaches sensory modalities as semiotically active factors and organizing principles in meaning-making. The focus will be on the special case where modalities mismatch in film – i.e., the soundtrack and visuals present contradictory meanings. The conflict can be characterized by the concept of synthesis that emerges in theories of Eisenstein, Barthes, Jakobson, Lotman, and cognitivists. The artistic functions of such synthesis will be discussed with the help of examples from selected feature films. In the first place, conflicting modalities are inspected in the light of Juri Lotman's theory of two incompatible, but still complementary languages that make up a mechanism for generating new information. In addition, the prospects of evaluating modality conflicts will be touched upon, dismissing synchrony and redundancy as the scale parameters, but acknowledging Lotman's model of space as a primary modeling system that is capable of representing semantic conflicts.

Keywords: film semiotics; multimodality; synthesis; creative conflict; Juri Lotman

1 Introduction

1.1 Creative synthesis

Semiotic study of multimodality is closely related to the concept of synthesis. Both are concerned with complex texts that simultaneously consist of different sign systems or require multiple sensory channels for processing. Roman Jakobson, who emphasized the need for the semiotic inquiry of multimodal phenomena (1971b [1964]: 339), called such texts *syncretic messages*.[1] Multimodal texts illustrate

1 "The study of communication must distinguish between homogeneous messages which use a single semiotic system and syncretic messages based on a combination or merger of different sign patterns" (Jakobson 1971c [1967]: 705).

***Corresponding author: Martin Oja**, University of Tartu, Tartu, Estonia, E-mail: martino@ut.ee

that meanings depend on relations: not only relations between texts and people, but between text's components or languages. Such relations can be made subject to various typologies, where *conflicting modalities* emerge as one relation type. It is not the most common, but nevertheless has a special importance for meaning-making.

In several theories of artistic expression, the idea of synthesis has an eminent position. Concerning film as an exemplary synthetic medium, the bimodal synthesis of auditory and visual modalities has been occasionally discussed. Nevertheless, what exactly are the characteristics of this synthesis, is not easy to specify. In film, for example, both modalities may carry similar meanings and thus support each other. For evolutionary reasons, most meaning-making and perceptual strategies are inclined towards such supportive process. But the contrary is also possible: auditory and visual modality may carry differing, even contradictory meanings. These are special cases with their specific purposes and functions. I refer to these cases as modality conflicts.

The origin of such processes can be traced back to biology. As Kalevi Kull claims, while logically congruent systems work like machines (Kull 2015: 616), the logical conflict is a requirement for meaning-making in all living systems. The situation of incompatibility facilitates choice between possible alternatives (Kull 2015: 618). At least partly, such ideas are congruent with Jakobsonian tradition, where incompatibility is seen as a creative element accompanying aesthetic function. A fundamental characteristic of an aesthetic text is that it allows the reader or viewer hold two opposite ideas in mind (Bennett 2021: 162) and, through that, promote a new hypothesis generation (Bennett 2021: 159).

Jakobson's thinking, in turn, had notable influence on Juri Lotman, who conceived a mechanism for creating new information. In this mechanism, two incompatible languages are juxtaposed. Lotman's understanding of the terms *language* and *artistic language* is remarkably general. For example, when Lotman writes, "a rhetorical effect arises when there is a conflict of signs relating to different registers" (Lotman 1990b: 51), he expresses the same idea.[2] This generality allows us to transfer the analogy to modalities, conceiving a similar mechanism where incompatibility between modalities generate new meanings.

The main objective of this article is to discuss the meaning-making potential of conflicting modalities in film, addressing the textual strategies and goals that conflicting modality as a semiotic practice serves. Analyzing various examples from feature films lets us examine how cognitive and textual aspects are related in such processes. Because of the limited scope of this article, I will abstain from touching the

2 Incompatible modalities can also be seen, under certain conditions, as codes. Two competing codes present the possibility for multiple interpretations.

problems of medium and multimediality, i.e., how different mediums affect the meaning of the invariant text. I will suggest a framework that is provisionally detached from the issues of media, but focuses on the interaction of *semiotic modes* and *sensory modalities*.

1.2 Modes and modalities

Semiotic modes can be understood as different forms of expressions and their combinations (Bateman et al. 2017: 44). *Modalities* are seen as organizing/interpreting principles that stem from psychological properties of interpreters. In the case of film, modalities characterize how meaning-making is processed according to the division into visual and auditory faculties.

In multimodal research, there are some notable problems related to the *modality* and *mode*. For example, theorists have difficulties in overcoming the divergent use of the terms. As Lars Elleström drew attention, "in … media studies and linguistics, 'multimodality' sometimes refers to the combination of text, image and sound, and sometimes to the combination of sense faculties (the auditory, the visual, the tactile and so forth)" (Elleström 2020: 41). Furthermore, as Bateman and Schmidt observe, "the precise nature of 'mode' in multimodality remains … unclear and a variety of descriptions circulate in the literature" (Bateman and Schmidt 2011: 75).

In a nutshell, the difficulties with *mode* largely originate from an eminent branch of multimodal reasearch, the sociosemiotics of Gunther Kress and his colleagues, who developed their theory on the basis of Michael Halliday's linguistics. Halliday's discussion of language's metafunctions led to the concept of *semiotic mode* as reflecting the use of materiality for the achievement of these metafunctions (Bateman et al. 2017: 49). Now the frame expanded from language to all kind of meaning-making activities, and the *mode* itself obtained a general meaning of *semiotic resource* (Boria and Tomalin 2020: 12, 13). As Kress explains: "every community has a range of resources for making meanings evident: speech, gesture, gaze, writing, and others; that is, the modes of social semiotic multimodal theory" (Kress 2020: 28). It is also important to notice that the notion of *semiotic mode* is not far from James Gibson's notion of *affordance* (Gibson 2015 [1979]: 119). As modes are shaped by the histories of their making in specific societies, they vary across cultures (Kress 2010: 130), and consequently, as units of research, are affected by a certain relativity.

Such relativity, in turn, brings along the lack of clarity and causes divergent interpretations by various theorists. It is well illustrated by the problem of submodes:

> The question of whether X is a mode or not is a question specific to a particular community. As laypersons we may regard visual image to be a mode, while a professional photographer will say that photography has rules and practices, elements and materiality quite different from that of painting and that the two are distinct modes. (Kress et al. 2000: 43)

As the previous example illustrates, it is difficult to subordinate the "level of detail" to a mutual agreement.

Yet another problem that theorists encounter is that aspects of mode as semiotic resource and modality as sensory modality get mixed up: "talking of 'modes' in terms of 'sensory channels' can be quite misleading. Sound is not just sound or the hearing of tones; it also gives information about space, hardness, distance and direction" (Bateman et al. 2017: 27). For the sake of clarity, I propose to acknowledge the difference between the meaning-making potentials that pertain to text (semiotic modes) and the meaning-making potentials that pertain to interpreter (sensory modalities).[3] Depending on that division, we can properly evaluate if *hardness, distance and direction* of the sound, which may be present in sound as mode, actually affects meaning-making on the level of modality.

Thus I suggest a framework where *semiotics of modes* is about the textual combinations and the potential of resources as affordances, the *semiotics of modalities* is about the meaningfulness of reception, cognitive processing and subjectivity. Paraphrasing Peirce, modes are oriented towards objects and sign vehicles, modalities towards interpretants. Consequently, modality as a unit in the psychology of multimodal perception – "sensory modalities are classically distinguished based on the type of physical stimulation that they are most sensitive to: light for vision, sound for hearing, skin pressure for touch, molecules in air for smell, etc." (Bertelson and De Gelder 2004: 141) – fits well into the proposed framework of multimodality, where a system of modes and a system of modalities can be seen as two complementary layers. The structure of sensory modalities could be seen as a foundational layer under the system of modes, which are "carried" by modalities. While the modes as semiotic resources are somewhat relative and their range is virtually unlimited, the modalities are concrete and their number is strictly limited. A framework that joins these two layers can therefore provide concreteness and flexibility at the same time.

As an immediate example of analysis using such framework, we can briefly look at a famous multimodal conflict in visual art. René Magritte's painting *The Treachery of Images* (1929) depicts a pipe and features a title "Ceci n'est pas une pipe." Here, the artwork employs only a single, visual modality, which brings along its practices of production and reception, and most importantly, requires certain

3 As mentioned beforehand, this is a purposefully simplified model where the problems of multimediality, i.e., the channel are omitted and, if necessary, expressed under modes and modalities.

cognitive mechanisms for processing it. While there is no *modality conflict* in that painting, there is *a conflict between modes*, and this allows us to call it a *multimodal conflict*. The painting features at least two eminent modes, which we can label "image" and "text."[4] As juxtaposed, they offer a possibility for semiotic synthesis. Also, we may recognize how visual modality sets certain restrictions and delegates certain affordances to the modes of image and text.

1.3 The structure of the article

The following study is divided into two parts. In the first, I will discuss several approaches to the idea that new meanings emerge in the intersection of two non-compatible systems, let them be codes, languages, modes or modalities. These approaches take place in the area where semiotics and film theory intertwine. The arrangement will be more or less chronological: first I will stop at Sergei Eisenstein's thoughts, then touch upon Roland Barthes' concept of *third meaning*, Roman Jakobson's notion of *syncretic message* and Juri Lotman's idea of two contradicting languages as a mechanism for creating new information. Lastly, I will look into cognitive film theory, which gives us an evolutionary perspective for understanding why the conflict between modalities is not common, but a special case with its purposes.

This all leaves the problem of defining conflict somewhat undone. Because of that, in the second part, I will discuss the notion of modality conflict and its dimensions. I propose four different grounds for modeling this conflict: first by a simple negative definition, then by the concepts of redundancy and synchrony. As the fourth, I consider Lotman's idea of space as a modelling system. First three possibilities are dismissed, while the potentiality of spatial model is acknowledged. Lastly, I will briefly discuss conflicting modalities in feature films, referring to the topic of *incongruent music* in recent film theory and also touching upon the usage of *internal diegetic sound* (IDS) as a meaning-making device. Like incongruent music, IDS relies upon modality conflict, but instead of music, it has (inner) verbal speech as the main mode carried by its auditory modality. Concludingly, in order to illustrate the functioning of IDS, a short interpretation of Hamlet's monologue in Laurence Olivier's film (1948) will be given.

4 In films, we also encounter a situation where two modalities carry the same mode. Textual (or verbal) mode can be represented by auditory modality (dialogue) and visual modality (subtitles) at the same time.

This is closely related to *amodal invariants* discussed by Taberham. These are common properties of objects that are represented in different modalities and can be perceived as carriers of the same meaning (Taberham 2013: 47).

2 Two modalities, two languages: meaning-making by synthesis

2.1 Sergei Eisenstein: montage of attractions and contrapuntal sound

2.1.1 Intellectual montage

The idea of juxtaposition was inherent in Eisenstein's philosophy of film-making. Eisenstein was a versatile person, trained as an architect and civil engineer, practiced as a theatre director, actor, graphic artist, film theorist, film director, writer, and teacher (Robertson 2009: 2). A synthetic approach seems naturally linked to such type of talent. As the influences on Eisenstein's thinking spanned from Wagner's *Gesamt-kunstwerk* to Dadaist art and the synaesthetic flow of consciousness in Joyce's *Ulysses*, at least two developments of the cinematic expression must be mentioned as major effects on filmmakers of this period. Parallel editing was developed, among others, by Porter and Griffith, and Kuleshov's famous experiments included *creative geography*, where a seemingly coherent space was built by details from various sources. These clearly expressed that cinema – still a young art – was much about the practice of synthesis.

The concept of *montage of attraction*, Eisenstein's well known contribution to the vocabulary of film theory, was initially inspired by the artistic activities outside of cinema. In general, Constructivist thinking, but more specifically, photomontage, circus, and the theatre of Meyerhold and his own, utilized the juxtaposition of active emotion-inciting moments (Goodwin 1993: 27, 28), leading the audience to a certain *ideological conclusion* (Eisenstein 1988 [1923]: 34), which is also an example of composite meaning. Still, this impact is not a result of a straightforward emotional programming with complete control. As a keen researcher of psychology and a part-time collaborator with Vygotski and Luria (Robertson 2009: 144), Eisenstein was well aware that audience's final reaction could remain somewhat open-ended.

Another link between Eisenstein and synthetic thinking is not technical but ideological. Eisenstein, known for his *dialectical* or *intellectual* montage depicting historic events (notably the 1905 uprising and October Revolution), was influenced by the Marxist-Leninist mindscape, which saw history as a complex entity, almost impossible to be fully manifested by artists. Thus, the historic condition calls for special expressive means, and the montage theory of Eisenstein tries to answer that call (Goodwin 1993: 83). This brings along synthesizing and condensing, as well as

meaning-making via powerful visual metaphors. For example, in *Strike* (1925), shots of workers attacked by cavalry are juxtaposed with bloody scenes from a slaughterhouse, creating a meaning of another level.

The seeds of the new meaning can be present in both sources separately. Joining them is a formal device that amplifies emotional impact and offers a rationalization or logical connection: *a* is like *b*. This example can be called a multimodal meaning-making only in the sense of modes as semiotic resources. While not being a modality conflict (as *Strike* is a silent film), it still reveals some issues relevant to various types of synthesis. *A* and *b*, more often than not, are not each others' (semantic) opposites. Rather they should share some common ground, i.e., something that allows connecting two spheres of meaning, seeing them in a single, integrated system with new functions and properties. In any case, the joining of *a* and *b* should not be seen as a simple operation of arithmetic as adding or multiplication. Instead, I propose that Eisenstein's compositions can be interpreted in the light of Juri Lotman's idea about two partly compatible languages, to which I turn later.

2.1.2 Contrapuntal sound and disassociation

Somewhat analogous to joining different shots in dialectical montage was assembling film with its soundtrack. Eisenstein didn't belong to the group of directors and critics who eshewed the nascent sound, afraid of its power to contaminate silent cinema's purity. On the contrary, he welcomed sound as an artistic possibility, taking an interest in what sound offered as creative means among other attractions. He coined the term *audiovisual cinema* to characterize how "sound film should work in terms of an interaction of music, sound and film as a unified form" (Robertson 2009: 13). The amount of Eisenstein's writing that refers to sound editing is extensive, and is well analyzed by Robert Robertson's superb monograph. Here, I will only turn to one text which has a central position in the discourse.

At the beginning of sound film era, 1928, Eisenstein, Pudovkin and Alexandrov published a short note called "Statement on Sound." These Soviet filmmakers stated theoretical premises for the evolution of sound film, delivering criticism towards Western cinema. They worried that sound coupled with visuals and representing dialogue in a lifelike manner may hinder the perfection of cinema as art, destroying the culture of montage. The particular object of their criticism was the commercial exploitation of sound, "in which sound-recording will proceed on a naturalistic level, exactly corresponding with the movement on the screen, and providing a certain "illusion" of talking people, of audible objects etc." The authors claimed that only the *contrapuntal use of sound* will afford new ways for the development of montage,

calling up for "a distinct non-synchronization with the visual images" (Eisenstein et al. 1977 [1928]: 257, 258).

The notion of *contrapuntal sound*, inspired by music theory, was introduced by Eisenstein, but later seldom used in the film discourse (e.g., Min Hong 2019; Richards 2008). Counterpoint, in the words of Robertson, himself a composer and filmmaker, is "the simultaneous and contrasting combination of two or more melodic lines or voices, held together by common motifs and harmonies" (Robertson 2009: 13). Fugue as a musical form that makes extensive use of the counterpoint can be seen as an example for Eisensteinian editing. Reflecting on the composition of *Alexander Nevsky* (1938), Eisenstein speaks of *vertical* and even *polyphonic* montage (Eisenstein 1957: 74, 78). This type of editing doesn't discard the horizontal progress, but operates equally on both axes. While a "track" develops horizontally, its elements are in continuous interplay with the elements of other tracks, establishing vertical relations. It is also important to note that all "tracks" need not only be in correspondence with each other, but also to the text as a whole. Vertical montage can work as a model for any multimodal text where tracks of different sensory modalities are simultaneously processed. There, semiotic modes have vertical relation with the modes of another modalities, constituting a certain vertical syntagmatics.

The notion of *contrapuntal* neither refers to a total synchrony nor a total disruption, but to the interplay between these principles; to a model where contrast and commonality are joined. This is important to understand while making sense of the *conflict* between modalities, which I will discuss more closely later. It is also interesting to observe that Eisenstein's call for disassociation between sound and image is close to Russian formalists's idea of *ostranenie* or defamiliarization. Viktor Shklovsky formulated this term to illustrate how art heightens perception and breaks automated responses (Stam et al. 1992: 10). As Laurent Jullier has exemplified, the concept of defamiliarization appears as an important locus for negotiating between two contrasting approaches in film theory: ecological and constructivist one (Jullier 2010: 139). Challenging both the automatic, evolution-shaped responses and the totality of cultural construction of meanings, it facilitates a synthesis between these standpoints.

Eisenstein's critique towards common filmic expression was largely pointed at the natural, lifelike style that dulls the viewer, instead of activating her. The dialogue between sound and image was called upon to generate disruptions where new information or a *third meaning* emerges. Some decades later, Roland Barthes approached Eisenstein's heritage through the framework of semiology. For him, the third meaning wasn't just a new emergent information in the shape of text-as-whole, but something more elusive, complicated, and enigmatic.

2.2 Roland Barthes and the "third meaning"

In a film scene,[5] Barthes distinguishes three levels of meaning: referential, symbolic, and the third one, which he eponymously calls *third meaning* (Barthes 1977: 52, 53). In first two, the code is notably established (in the first more firmly than in the second), but in the third layer, signifiers remain without a clear referent; the third level is open-ended, disclosed to potentialities. Each next level demands a more developed, and I guess, more abstract semiotics. The third, also called *obtuse meaning* is a signifier "without a signified, hence the difficulty in naming it" (Barthes 1977: 61). It is something that emerges from the fabric of the text itself, from the details, often not intentioned by the author. As such, the reading of the third level of meaning – reading in general sense as reading an image or film – is erratic, and the reader could not be completely sure if that level is even justified.

Barthes sees the third meaning as *a new, rare practice* which goes against major signification practices; a luxury without rational output, something that belongs more to the future than present (Barthes 1977: 62), and is an epitome of counter-narrative (Barthes 1977: 63). Although the third meaning does not directly relate to the emergence of new information through the interaction of different sign systems, the idea of it as something out-of-ordinary, something that subverts the story (Barthes 1977: 64) goes well together with the observation that neither conflicting modalities do not belong to everyday practice of film-making. To this topic, I will more closely return below.

The third meaning, while active, has influence on other signification levels. If we develop Barthes' idea further, we can imagine a situation where this "obtuse" meaning takes over the *referential* and *symbolic*. The possible mechanism behind this process could also be envisioned, and this is, for the semiotics of multimodality, enlightening. Namely, on the referential or symbolic level, several modes and modalities may compete with each other and demand a certain solution for the fixation of meaning. In that case, the third level could incite the reader or the viewer either to (1) choose between variants or (2) synthesize a meaning on the basis of conflicting material. It is highly probable that this binary choice between choosing and synthesizing would depend on the extratextual and contextual factors.

Perhaps the most revealing to our topic is the section where Barthes picks up a thought that illustrates the multimodal synthesis (Barthes 1977: 62). It originates from Eisenstein's discussion of color while working at *Ivan the Terrible* (Goodwin 1993: 204): "the creaking of a boot seen on the screen is not art. Art begins the moment

5 More exactly, he looks at the stills from Eisenstein's *Ivan the Terrible* (1944). This is not a multimodal semiotic act, *per se*, but the conclusions he makes about the "third meaning" can also apply to the case where several modalities interact.

when the creaking of a boot on the soundtrack is related to a different visual image and thereby stimulates corresponding associations" (Eisenstein 1961: 84). Here, Barthes refers to the synthesis that retains the multiplicity of possible meanings: "multi-layering of meanings which always lets the previous meaning continue, as in a geological formation, saying the opposite without giving up the contrary" (Barthes 1977: 58). This again brings forth the problem of interpretation. In the textual environment where multiple signs function together, the third meaning assumes a higher degree of interpretational freedom by viewer who has to conduct synthesizing operations.

Influence of the Saussurean semiology that holds the relation of signifier and signified in the central position, leads to the problem of text's dual nature (that is also frequently discussed by Lotman). A text can be seen both as a combination of independent signs and a whole sign itself, having its place in the discourse or semiosphere. This duality has its implications on Barthes' *third meaning* and the meaning generated by a multi-modal conflict, as well. This can be called the "problem of the unit." As Barthes finds it difficult to delineate research objects on his third level, the synthesis between modalities forces us to rethink what units are relevant. In the preliminary phase of methodology-building, modalities itself (as auditory and visual in the case of film) can be treated as research units, and as I suggest, Roman Jakobson has more or less followed this idea.

2.3 Roman Jakobson: dominant in syncretic messages

Although Jakobson assigned verbal messages the primary role in communication, he was keenly interested in other communication systems and their mutual influence, as well. While multimodality studies based on Kress's sociosemiotics have mostly steered clear of sensory modalities and manifested *mode as a semiotic resource* for the main research unit, exactly Roman Jakobson's work allows us recognize modalities as a topic of interest for semiotics.[6] In "Language in relation to other communication systems," Jakobson acknowledges senses as the preconditions for any type of signification:

> All five external senses carry semiotic functions in human society … Within the systems of auditory signs never space but only time acts as a structural factor, namely, time in its two axes, sequence and simultaneity; the structuration of visual signantia necessarily involves space and can be either abstracted from time. (Jakobson 1971c [1967]: 701)

6 Also, it is interesting to note parallels with Émile Benveniste's standpoint: "The mode of operation is the manner in which the system acts, more particularly the sense (sight, hearing, etc.) to which it is directed" (Benveniste 1981: 11). Under what Benveniste calls *translinguistics* or "second generation semiology" (Benveniste 1981: 21) we can allocate the research of multimodal sign systems.

Visual and auditory modalities constitute a duality in relation to the physical structure of the world: one relates to the space and the other to the time. Both parameters are nevertheless interrelated and transposable, e.g., transfer of meaning from oral language into written language is thus a transposition from the dimension of time into the dimension of space (Jakobson 1971c: 706). The same process is conceivable in the terms of code. Time and space can be seen as two supercodes, while the culture has invented means for the mutual transposition between these codes. These means do not guarantee a perfect transposition, but the incompatibility functions in the center of a model where two partly translatable languages have a tendency to generate new information.

The cultural means of such a transposition are signs: Jakobson acknowledges correspondence between the famous sign types of Peirce and sensory modalities – "the prevalence of icons among purely spatial, visual signs and the predominance of symbols among purely temporal, auditory signs" (Jakobson 1971c [1967]: 701). Inside both modalities, there can be found a typology of signs with their corresponding properties. In his essay "Visual and auditory signs," Jakobson writes:

> In our everyday experience the discriminability of visual indexes is much higher, and their use much wider, than the discernment and utilization of auditory indexes. Likewise, auditory icons, i.e., imitations of natural sounds, are poorly recognized and scarcely utilized … the supremacy of sight over hearing in our cultural life is valid only for indexes or icons, and not for symbols. (Jakobson 1971a [1963]: 335)

This naturally poses the question of the *dominant* as an important part of the framework. Modalities' relation to dominant as the main determining component of artwork's meaning and organizer of its structure (e.g., Jakobson 1981 [1935]: 752) needs a more comprehensive discussion than this article can provide. In short, psychologists have traditionally held an assumption about visual modality as dominant, but latest research have demonstrated that attention in different modalities is not independent (Shams and Kim 2010: 272) and the weighting of visual cues can be affected how consistent is the visual cue with the non-visual cues (Shams and Kim 2010: 277). It is also important to note that the nervous system has a constant task to figure out which sensory signals are caused by the same object and should they be combined (Shams and Kim 2010: 279). In other words, we tend to perceive our environment[7] as a whole, building a unified model to represent it in our consciousness.

Therefore, we cannot state that a certain modality is dominant *per se*, but have to consider the characteristics of the specific semiosis. For instance, experiments have indicated that for detecting the temporal resolution of events, audition is superior

7 Here I mostly agree with cognitive film theorists that the perceptual mechanism for cultural texts is based on the processes we use to perceive our natural environment.

over vision (Bertelson and De Gelder 2004: 149). This is in accordance with Jakobson, who asserts that the spatial dimension takes priority for visual signs and the temporal one for auditory signs: "A complex visual sign involves a series of simultaneous constituents, while a complex auditory sign consists, as a rule, of serial successive constituents. Chords, polyphony, and orchestration are manifestations of simultaneity in music" (Jakobson 1971a [1963]: 336). Jakobson compares paintings with non-specified auditory material (most likely speech or song which has verbal components), thinking that "when the observer arrives at the simultaneous synthesis of a contemplated painting, the painting as a whole remains before his eyes, but when the listener reaches a synthesis of what he has heard, the phonemes have in fact already vanished" (Jakobson 1971b [1964]: 344). Here, the key question is, how the specifics of short-term memory relate to the different types of world (or text) modelling. However, this topic would also require a treatment of its own.

Concluding with Jakobson's thoughts on multimodality, regardless of the solid foundation he offers to the framework of modality research, I have to point out a possible risk in it. Following Jakobson too closely can tempt us mixing up the opposition *simultaneous/sequential* with another opposition, *static/dynamic*. Jakobson himself was seemingly somewhat affected by this confusion. When he approached painting as an object for simultaneous synthesis, he did not consider the necessary activity of "reading" the picture in steps, as eye tracking studies have later confirmed (see, e.g., Smith 2013: 167). Thus in the case of image perception, the simultaneous and sequential types of synthesis work together.

Nevertheless, we can identify a difference between static texts (painting) and dynamic ones (music, film) and also take notice how the time we take processing texts relates to the "duration" of the texts themselves: are the temporal borders fixed (as in film or recorded piece of music), or undefined (as in still image or written text). Syncretic messages, as Jakobson addressed multimodal texts, are dynamic and have fixed duration, as a rule. Next, we will return to the semiotic potential of time and space regarding Juri Lotman, who, in his later period of thinking, had the idea of space as a language, functioning as an alternative primary modeling system.

2.4 Juri Lotman: multiple languages, new meanings

For Juri Lotman, the notion of *language* is notably general, and more extensive than the traditional verbal concept, which only refers to natural languages. In his last monograph *Culture and Explosion*, Lotman defines language as a flexible code with history or cultural memory built into it (Lotman 2009 [1992]: 4). This allows us to apply Lotman's language-centered model to a wide range of semiotic systems, including the system of sensory modalities. Regarding a modality in the position of

code we can recognize the animal species' genetic features[8] that concern the processing of a certain modality; in the position of cultural memory we can see the semiotic practices that have been developed by culture over time and are pertinent to that specific modality: e.g., for the auditory modality, the practices of making sounds and listening them, and also technological means to facilitate these practices.

The notion of *artistic language* enables us orient even better. For Lotman, art is an activity of meaning-making that contains the idea of an inherent multiplicity and dialogue. Thus, an artistic work contains:

> a chorus of simultaneously speaking languages. The possible relationships among them are various: any one may occupy a dominant position, imposing its modeling system on all the others, or the "languages" may be distinct from each other or even mutually contradict each other, forming a contrapuntal construction. (Lotman 1990a: 211)

Here, in this passage, we hear the echoes of Mikhail Bakhtin, and also of Jakobson, who had a strong influence on the development of Lotman's ideas. Nevertheless, a notable inference can be made from the quotation above: the hierarchy of different languages in a single artistic system is not fixed, but *dynamic*. Consequently, different languages or modalities can occupy the dominant position. In an artistic system, e.g., in film, visual and auditory modalities can regularly change each other's place as the organizing principle of meaning-making. The same goes for semiotic modes, but as a rule, in a more frequent time-scale.

Thus, an event of contradictory languages or a *modality conflict* does not automatically imply a dominance of a certain modality. It rather marks the end of one domination regime and leads to an open-ended situation where there are two or more contending variants of meaning; the viewer has to choose between them, thus participating more actively in the semiosis. For Lotman, as for Eisenstein, this is the core artistic mechanism. Inherent to that mechanism is dialogicity. "Dialogue presupposes asymmetry, and asymmetry is to be seen first in the difference between the semiotic structures (languages) which the participants in the dialogue use" (Lotman 1990b: 143). I suggest that two applications of the *dialogue* are equally relevant here. On one hand, the viewer is in dialogue with a film; on the other, multiple languages of the same text (i.e., the modalities and modes of a film) are in mutual dialogue with each other. Thus, the viewer has a position of an active moderator in the dialogue between languages or modalities.

---

8 It is interesting to notice here a converging point between cultural semiotics and biosemiotics. Biosemioticians would ask how the system of modalities and system of modes relate to different species; in other words, it's a problem of culture in alloanimals. It would be easy to apply a dualistic partition, reserving *modes* only for humans, and *modalities* for all species, humans included. Yet, such an opposition is quickly proved fallible when describing *semiotic resources* in alloanimals' behavior.

For the activity-inducing mechanism of a conflict, it is hard to find a more pertinent observation than Lotman makes in *The Structure of the Artistic Text*: "juxtaposed units that are incompatible in one system force the reader to construct an additional structure in which the incompatibility is eliminated" (Lotman 1977: 283). This implies for the need of unity and understanding; something that motivates a reader in her action, and even prompts her to enter the dialogue. Consequently, we must also face the question of *initial condition* for the dialogue. Lotman calls it the *dialogic* or *semiotic situation* (Lotman 1990b: 143, 144), implying a will to search for a common language. This, in turn, presumes a pre-established contact, and an area of partial intersection of lingual spaces (Lotman 2009 [1992]: 5). Hence, the languages in a dialogic system cannot be too similar or too different; the components of known and unknown are both required.

Concerning the problem of modality conflict in audiovisual texts, Lotman's distinction between discrete and continuous types of languages is also relevant. One of the key works concerning this topic is "Rhetoric as a mechanism of meaning-generation." Here, Lotman discusses tropes as the central figures of rhetoric, indicating that tropes are created in the contact point of two languages. As such, the schema offers a model for the creative consciousness itself (Lotman 1990b: 44). In this bi- or multilingual system, there is a tendency of conflict between discrete and continuous types of coding, which are mutually incommensurable, and therefore, intranslatable. Nevertheless, the struggle to translate generates new information on the level of metalanguage (Lotman 1990b: 36, 37). In the context of this article, the question is, can multimodal conflicts be modeled on the basis of mismatch between discrete and non-discrete coding types?

First, I would point out that a multimodal text as a whole tends to rely on non-discrete logic, but may contain discrete languages inside of its structure. An example of this is verbal text integrated into film, or as Lotman puts it, "cinematic metaphor is built up by relating the shot to natural language discourse, and so the mechanism of discreteness is brought right into the structure of the cinema-metaphor" (Lotman 1990b: 38). On the other hand, the answer would largely depend on whether the viewer consciously distinguishes specific narrative units or, instead, handles the meanings as a continuous process, e.g., emotional flow that fluently changes from a state to another. Both sides have their advantages and can be mutually complementary.

Therefore, it is not justified to call one modality discrete and the other non-discrete, *per se*, even though for Jakobson, the visual modality relied on *spatial*, the auditory modality on *temporal modeling*. I suggest that *auditory/visual* opposition is not necessarily in correlation with *discrete*/*non-discrete* opposition. Nevertheless, if we turn to *modes* as semiotic resources, discrete and non-discrete principles can be seen as distinguishable. It has to be noted that while Lotman juxtaposes *verbal* and

visual as the representatives of discrete and non-discrete systems, this is incommensurable with our system of modes and modalities. Thus, both models must be brought into mutual correspondence: *verbal* is a *mode*, but *visual* a modality. In other words, *verbal* as a mode can be carried by both modalities, *visual* and *auditory*. That we should leave dualist models and turn towards more integrated, body-centered and evolutionary descriptions, is a proposition forwarded by cognitive film theory.

2.5 Cognitive approach and evolutionary perspectives

2.5.1 Embodied simulation and film viewing

While Roland Barthes considered the *third meaning* the epitome of the *counter-narrative*, it correlates surprisingly well with the cognitivist stance on narrative cinema as embodied simulation. Following the cognitivist paradigm, we can associate the synthetic, emergent meanings with the situation where the conventional narration stops and the viewer is called up to compose her own meanings. I have discussed this mechanism before (Oja 2014: 87) as a device inherent to art films, where breaks in narrative activate viewers' subjective meaning-making.

The concept of *embodied simulation* or mental simulation (e.g., Grodal 2009: 150) rises from the integrated perceptual model: "with the support of contemporary cognitive neuroscience, it is possible to formulate a new perceptual model in which action, perception, and cognition are closely integrated" (Gallese and Guerra 2015: 151). This somewhat holistic approach was anticipated by the phenomenology of Maurice Merleau-Ponty and more specifically by Vivian Sobchack. Sobchack's wordplay with *cinesthetic* refers to viewing as a synaesthesia-like experience (Sobchack 2004: 67), where body functions as an integrated organ of reception and meaning-making. A thought-provoking angle on the embodiedness is also provided by Laura Marks, who discusses the body as a source of memory, the cultural practices of operating that memory, and certain cross-modal relations as *haptic visuality*, where visuality functions like the sense of touch (Marks 2000: 22), i.e., when seeing someone touch a thing or another person, the mirror neuron system activates the processing circuit of touch in spectator's body.

As Gallese and Guerra point out, there has been a paradigm change in cognitivist film theory as well. While classic cognitivism insisted on a modular concept of mind and applied computer-like analogies to human consciousness, neuroscience has demonstrated that human senses and action schemata, i.e., our perceptional and motor systems are more integrated than previously thought (Gallese and Guerra 2015: 151, 156). This leads to an updated understanding of perception, which is, in a

sense, always multi-modal. The concept of embodied simulation[9] depends on the fact that perception of other person's actions and emotions is based on the activation of the same neuronal, and most importantly, motor circuits. This enables us recognize others' experiences directly, without conceptualization or symbolic mediation. This pre-reflective mirroring enables us simulate the states of others and also understand fiction that employs visual modality.

Sensory processing is not only done by synthesizing visual and auditory information in high level brain circuits or convergence areas, but also on the lower level where other senses (as touch and proprioception) are to some extent cross-activated. Filmmakers, more or less knowingly, use various techniques of sensory immersion that guarantee embodied simulation (Gallese and Guerra 2015: 54) such as camera movement which is similar to the body's natural progress in its environment, or close-ups that incite stronger engagement with film (Gallese and Guerra 2015: 91, 111). Normally, in the transformation and processing of multimodal signals, it is crucial to locate the sources of information and represent them coherently. Therefore, the coding (encoding on the side of film-makers, decoding on the side of viewers) has to be organized accordingly. In embodied simulation, multiple sensory modalities should work in synchrony, both for the processing of fiction and orienting oneself in the environment, or so-called "real world." Thus, synchrony is a default practice, and the conflict a special case.

2.5.2 Modality conflict is an exception, not a rule

The *suspension of disbelief* is a popular phrase in film discourse, having its roots both in nineteenth-century aesthetic philosophy and twentieth-century psychoanalysis. The idea is, while encountering fiction in its various forms, the viewer or reader must make an effort to forget the "real world" with its rules and start to believe in features of the fictive world.

Cognitive film theorist Torben Grodal is critical of the term. He claims that suspension of belief is needed only so much that the viewing does not produce full-scale illusions; otherwise, even if we watch fictional films, the seeing is believing, because the believing is the default mode and actually *disbelieving* demands special effort (Grodal 2009: 154). For this, and for the previously indicated reasons, there is a strong incentive for a filmmaker to create modally coherent storyworlds.

---

9 *Embodied simulation* also reflects James Gibson's classic concept of *affordance*: instead of objects' qualities, we perceive affordances or the things we can do with these objects (Gibson 2015 [1979]: 126). Consequently, perceptions trigger various action schemata in us and the meaning-making is decisively influenced by the characteristics of our bodies (or, why not, our bodyminds). This, in turn, is in compliance with Jakob von Uexküll's concept of Umwelt and Maurice Merleau-Ponty's phenomenology of perception.

The inclination to believe rather than not to believe, the search and attention towards truthful cues is largely determined by evolutionary factors. As Joseph Anderson states, the illusion of film's reality depends even more heavily on sound than on image. Audiences are more tolerant to picture glitches;[10] if something is interrupted or unnatural on the soundtrack, it induces an instant feeling that something is wrong (Anderson 1996: 80). Overall, perception is an information-gathering activity, and survival has demanded of perception veridicality (Jullier 2010: 122). We cannot afford to be too wrong about what is happening around us. When information occurs in multiple modalities simultaneosly, we start the comparison, "an active search for cross-modal confirmation" (Anderson 1996: 82, 89). Here, a special type of redundancy is characteristic to *normal* situations; we can call it the case of *mutually supporting* modalities. The decrease of that redundancy does not automatically induce the conflict, but I will turn to this question more closely in the next section.[11]

Below, I will also discuss the basis for the evaluation of modality conflicts. According to Anderson, synchrony serves as a linkage mechanism at very low levels of a perceptual system. That is, independent but simultaneously appearing features of complex systems are represented by synchronous firing of the cells. In this process, the features are usually bound together (Anderson 1996: 83) into larger meaningful systems or complexes. This is, in my opinion, a significant aspect of semiosis; the origin of the semantics of compound entities. As such, the notion of synchrony would refer only to the proximity in time, but I suggest the analogical mechanism of correlations can be described also on the basis of spatial characteristics.

3 What is a *modality conflict*? Dimensions and measures

3.1 Dimensions of multimodal conflict

While Eisenstein criticized reality-mimicking sound in commercial cinema, the creative approach he suggested would set the information from visual and auditory modalities against each other. The viewer could then be brought out of automatism,

10 The reason behind this is that we blink frequently, even without awareness (Anderson 1996: 80), and our brains have developed mechanisms of coping with these gaps of visual continuity.

11 Similar comparison would arise when we encounter conflicting *modes* inside a single modality. While during modality conflict, the feeling of oddity and alarm would rise, I propose in mode conflict we tend to have a feeling of extraordinariness and semantic confusion. Seeing the sun and rain simultaneously (where a rainbow emerges as a natural *third meaning*) has an extensive reflection in folklore all around the world.

inciting a more active reception where novel meanings are constructed. The premise for such a synthesis is a recognition that one needs to find a common ground for divergent modalities, and consequently, interpreting conflicting components in each other's terms. This can be understood, at least metaphorically, as translation. For modality conflicts, such a translation should be intra-systemic, i.e., between the text's components. Still, the need for translation may arise for another reasons, modality conflict is not the only factor that may hinder semiosis. If we approach the conflict as some type of difference, solving the conflict manifests itself as a wish to grasp the meaning of that difference.

From the previous section, we brought along the negative definition of multimodal conflict. Namely, the conflict emerges when mutual support between modalities is suspended and different modalities start to express divergent meanings. Although the negative definition is instrumental, the notion of *mutual support* needs a closer examination, and also some attention has to be payed to the *positive definition* of conflict: what is it exactly? Is it a contradiction, negation or incompatibility, divergence or difference? Can we detect it due to intranslatability or partial translatability, or due to translatability with some quirks and difficulties? If the answers are somewhat affected by the relativity of signification, as we may expect, how then to model the scope of conflict? First I will discuss the possibilities to define conflict on the basis of the degree of redundancy and synchrony, then by the dimensions of time and space.

3.2 Redundancy

The notion of redundancy travelled towards the semiotic discourse by multiple routes: most significantly via linguistics, and then via communication engineering, which, through cybernetics, inspired Roman Jakobson as well as Tartu-Moscow school in the 1960s. From the standpoint of a communicative act, the function of redundancy is clear-cut: to guarantee that a message reaches the addressee in the same form as it was at the time of departure. Analogic communication systems were prone to introducing noise: imperfections in the channel could modify the end result. Redundancy was seen, mostly by Shannon (1948) in his seminal article "A Mathematical Theory of Communication," as a counter-measure against noise. The most obvious way to introduce redundancy into text is to double its elements or repeat the whole message.[12] A good example of multimodal redundancy is a traffic light that

12 James Gleick, in his book *The Information*, recounts a story about a lawsuit against Western Union Telegraph Company in 1887. The firm was defended by a fine print on the telegraph blank saying that the company shall not be liable for mistakes in an *unrepeated* message (Gleick 2011: 158).

emits sound signals in addition to its color scheme. To be exact, this is the redundancy between sensory modalities. In traffic lights, there can also be another type of redundancy which uses modes: a *color* plus human *figure* (either standing or walking) are two modes that both rely upon the visual modality.

For an artistic message, the theory of communication ceases to be mathematical; in best case it remains statistical. A similar case is with situations where certain events in an environment are being interpreted as informational. Talking about such events (e.g., a hunter hears something rumble in the undergrowth), we cannot consider them messages *per se*, but rather types of semiosis that involve symptomatic or "natural" meanings (see, e.g., Forceville 2020: 15). Therefore, in such occasions there are no intentional senders but addressees interpreting some materiality or some pattern as information; they may assign such patterns a status of message *post hoc*.

When living agencies receive information, they bring along their subjective *Umwelts*. The *Umwelts* function as personal contexts that can be different in sender and receiver. Despite the possible existence of the mutually verified code, the message is actually no more *received* than *interpreted*. Consequently, the rate of redundancy cannot be precisely measured. Still, when considering evolutionary incentives to get coherent information, the concept of redundancy has some explanatory power. For the sake of survival, hearing a dangerous-sounding noise, the same hunter benefits from visually locating the source of that noise. In such case, conflicting inputs from different sensory channels can lead to unpleasant results.

In complex multimodal texts, e.g., film, the notion of redundancy is substantially more problematic. In 1980, Rick Altman observed retrospectively that many film theorists have considered the soundtrack of classical narrative films (in contrast to experimental, avant-garde films) redundant, only intensifying the sense of reality provided by the image. Altman, however, noticed that in classical narratives usually the soundtrack makes *image* redundant, not vice versa: the soundtrack is like a ventriloquist who makes image his dummy, creating an illusion that the words are produced by the image. In such a case, talking about redundancy between image and sound is inadequate. Without the dialogue, the images are ambiguous, incomplete, and undetermined (Altman 1980: 68, 69).

As Michel Chion indicates, when sound adds meaning to image, this meaning often seems to emanate from the image alone. It leads us to project additional value onto the image. Nevertheless, the feeling of redundancy between sound and image is an illusion. Even a most common example of a filmed dialogue between two interlocutors is far from redundant. Faces and gestures of characters, costumes, details of location, etc., cannot be ascertained from the sound alone (Chion 2016: 152, 153). Consequently, "you cannot study a film's sound separately from its image and vice versa. In fact, their combination produces something entirely specific and novel,

analogous to a chord or interval in music" (Chion 2016: 161). These are strong arguments against modality redundancy even in mainstream films, not speaking of experimental films where modalities could be purposefully juxtaposed. Even in classical narrative cinema, the meanings emerging from the simultaneous effects of visual and auditory modalities are compositional, if not synthetical. For such reasons, the level of redundancy is not a sufficient measure for the scope of multimodal conflict.

3.3 Synchrony

Let us briefly return to the evolutionary concern. If, according to a simple provisional model, the simultaneous appearance of visual and auditory information leads to a coherent representation of an event, does asynchrony, on the other hand, automatically lead to the conclusion that something is "wrong" in that event, or, as another option, that there is an error in the perceptual system? Moreover, could we model modality conflicts on the basis of asynchrony, e.g., see a stronger conflict in the case of a more extensive asynchrony? I propose that such a model has serious shortcomings. Below, I will highlight three main spheres of problems.

The first sphere concerns natural meanings, where asynchronies are uncommon, but still observable. For example, viewing an event from a long distance, asynchrony is introduced by different speeds of light and sound. Yet, seeing a person hacking wood or firing a gun from afar, the repetitive nature of the activity or recurring experience of the event over time (as seeing lightning and hearing thunderclaps) suggests a pattern and regularity. In such cases, the meaning that emerges in the observer is not so much related to conflict but to the *modification* of perception; the knowledge of the delay can be quickly automated and integrated into the system of habituative or cultural codes. Then paradoxically, such asynchronies can be interpreted as synchronies on a deeper level.[13]

A second sphere of problems relates to artistic, purposefully created texts, and here we encounter a similar mechanism: asynchrony on a lower textual level helps to create synchrony on the higher level. This issue is well highlighted by two film editing techniques, namely, J-cut and L-cut. There is yet no comprehensive discussion of these techniques in academic literature, although on the basis of Frierson,

13 The other side of the problem is, when there is no pattern nor an observable regularity between the occurrences, it will be hard to interpret asynchronous auditory and visual information as a single event at all. Even if auditory and visual stimuli emanate from the same source in the same location, but occur outside of such code as mentioned above, they are with high probability interpreted as two independent events. For that reason, asynchrony may easily remain unnoticed and compositional meaning is not made.

definitions can be brought out. In J-cut "the sound for the incoming shot precedes the picture; an edit 'prelaps' the audio of the incoming shot into the outgoing scene. In contrast, in L-cut the sound of the outgoing shot continues after the picture ends; an edit 'postlaps' the audio of the outgoing shot into the incoming shot" (Frierson 2018: 319).

These techniques have been named by the iconic principle. When an editor looks at the computer screen where her workspace is represented, various tracks of visual and auditory material are displayed one above another. By convention, video tracks are displayed firsthand, i.e., above, and audio tracks below the visual. For that reason, the "overhangs" in the cutting points resemble the figures of J and L respectively.

For the theory of multimodal meaning-making, an extensive analysis of L-cut and J-cut would be enlightening. Here, I can only draw attention to the fact that these cuts can work both ways: as devices that create coherence on a higher level, and in special cases, as providers of specific psychological effects. Broadly, they belong to the system of *continuity editing* (CE), a filmmaking practice that provides narrative continuity. The purpose of CE is to disguise how the story is told. With its conventions, CE strives to back up a seamless, spatially and temporally coherent narrative (Hayward 2000: 74). Frequently, L-cut and J-cut, although seemingly initiating a modality conflict, solicit smooth transitions between scenes and segments, blending a shot into the next one and suggesting a connection between two timespaces.[14]

Lastly, the third sphere of problems entails the most significant argument for dismissing synchrony/asynchrony axis as a measure for multimodal conflicts. Paradoxically, most modality conflicts are created on the basis of synchrony, not asynchrony. Juxtaposition of modalities requires a simultaneous presentation of multiple textual elements. I will point out two quick examples on the basis of David Bordwell and his colleagues, who examine how filmmakers connect sounds to images.

Concerning the conflicting modalities, *rhythm* and *fidelity* are most relevant sound properties in the typology of Bordwell and his co-authors. They indicate a possibility for contrast between the rhythms of image and sound. Referring to

14 An inventive example of a J-cut can be found in Anthony Minghella's *The Talented Mr. Ripley* (1999) around 00:24:30, edited by Walter Murch. The scene follows Tom Ripley (Matt Damon), who has sneaked into Dickie Greenleaf's (Jude Law) room. In front of the mirror, he inspects Dickie's trinkets, trying on his wristwatch. A suspense is rising, inciting a question: will Tom be discovered by Dickie or his girlfriend? Then a J-cut: the sudden voice of Greenleaf is brought into the scene. For a couple of seconds, the viewer is tricked into a false realization that Ripley is discovered by Dickie. However, when the visual track is also cut, it turns out that another scene has started and the sound actually belongs to the new scene. Here, the modality conflict playfully undermines the paradigm of CE for a brief time.

Tarantino's *Reservoir Dogs*, they observe: "one of the … options is to edit dialogue shots in ways that cut against natural speech rhythms … If the source of sound is primarily offscreen, the filmmaker can utilize the behavior of onscreen figures to create an expressive counterrhythm" (Bordwell et al. 2019: 283). This technique partly follows Eisenstein's idea of contrapuntal montage: while Eisenstein's main purpose was creating intellectual meanings by the synthesis of two propositions, juxtaposition of image rhythm with sound rhythm can generate more bodily, affective meanings.

The category of *fidelity* is even more illustrating how multimodal conflict is based on synchrony and simultaneity. Fidelity refers to the extent to which the sound is faithtul to its source as we conceive it. Bordwell et al. point out how "unfaithful" sounds can produce estranging or comic effects:

> if a film shows us a barking dog and we hear a barking noise, that sound is faithful to its source; the sound maintains fidelity. But if the image of the barking dog is accompanied by the sound of a cat meowing, there enters a disparity between sound and image – a lack of fidelity. (Bordwell et al. 2019: 284)

These examples demonstrate that asynchrony seldom promotes a modality conflict; in most cases, synchrony is exactly a requirement for such constructions. Occasionally, as in L-cut and J-cut, a modality conflict on a lower level (scene or segment) facilitates continuity on a higher level (text or discourse): here the asynchrony is rather illusory. As was revealed by the example of unfaithful sound, juxtaposition of modalities generally requires simultaneity, and if a conflict is generated, it is not temporal but semantic.

3.4 Space as a modeling system

If the essence of multimodal conflict is a semantic divergence, the main question is, how do we measure, quantify, and model this divergence? As we saw, the degrees of redundancy and synchrony appeared unfit as the indicators. Looking for an alternative, I will shortly return to Juri Lotman's idea of space as an alternative primary modeling system.

In the semiotics of Tartu-Moscow school (TMS), the idea of primary and secondary modeling systems has an important place. Due to the structuralist background of TMS, the language[15] (in the sense of natural language) was considered

15 As we noticed above, Lotman's notion of language extended to various sign systems, including non-verbal ones. Thus he employed a dual understanding of language, a narrow-verbal, and wide-metaphoric.

primary, and literature with other systems of artistic expression, as secondary or supralinguistic systems that rely on language (Lotman et al. 2013 [1973]: 72). This idea was remarkably criticized by Thomas Sebeok, who took into account animal communication, pointing out that in human species, language was adapted not until a certain phase of evolution (Sebeok 1991: 334, 335). Therefore, language is established upon the more basic systems of cognition and communication; in many situations, "thinking without language" is primary. Consequently, a shift in the forementioned typology was called upon: language comes only as secondary modeling system, and supralinguistic systems have tertiary position.

Still, it is important not to dismiss TMS's and Lotman's idea about the primacy of language as naive or non-informed in biological sciences. I propose that Lotman's interest in nonverbal sign systems gradually expanded, embracing visual arts, social rituals, film, etc. For example, in his latest monograph *Culture and Explosion*, Lotman took interest in the semiotics of animal behavior and its ritualistic aspects (e.g., Lotman 2009 [1992]: 29). Thus, claiming language primary is a conscious academic self-positioning in a specific context. This is supported by Lotman's claim that *space* can also be seen as a primary modeling system. I see this as an indirect answer to Sebeok's criticism.

Seeing space as an active entity, not as a sterile background, is related to life sciences and especially Vernadsky's understanding of biosphere, which inspired Lotman's concept of semiosphere (Lotman 1990b: 125), the semiotic space that functions as the precondition for any semiotic activity (Lotman 1990b: 123). But perhaps most intriguing here is the transformation of the notion of *space* from the semiotic background to the *language* itself, or at least to a certain materiality of the language:

> The language of spatial relations … is not the only means of artistic modeling, but it is important, since it belongs to the primary and basic. Even temporal modeling often represents a secondary structure on the spatial language. (Lotman 1990a: 239)

Here, the possibility of space as an alternative or complementary primary modeling system is clearly emphasized.

In his analyses of culture, Lotman takes notice how the relations between textual elements can be in correlation with relations between components in a spatial model: "…the structure of the space of a text becomes a model of the structure of the space of the universe, and the internal syntagmatics of the elements within a text becomes the language of spatial modeling" (Lotman 1977: 217). So, as Peeter Torop observes, the textual space for Lotman is not simply the graphically fixed sphere of information, but an interpretational space (Torop 2022: 582). It applies both to verbal texts – e.g., discussing Gogol's prose, Lotman shows how characters' ethical positions

can be expressed with spatial models (Lotman 1990a: 202), as well as to non-verbal discourses, e.g., city planning: "Petersburg can rightly be considered to be … a place where semiotic models were embodied in architectural and geographical reality" (Lotman 1990b: 202).

As the spatial modeling system is comprehensive and rather universal, I propose it can be applied to the conflicts between modalities and semantic incongruences in them. The parameter of distance in spatial models can be set into correspondence with semantic distance in multimodal conflicts. Although the notion of *semantic distance* is in the danger of remaining somewhat relative and subjective, an attempt to anchor it in empirical evidence can be envisioned. Considering the architecture of the brain, processing different modalities can be seen as corresponding to the distances in a spatial model; the metaphoric spatial model can be seen realized in the neuronal organization. Moreover, spatial distances require energy to overcome, and in the case of conflicting modalities, additional energy is needed for the intrinsic translation and synthesis. In a similar way, creation and retrieval of memories is energy-consuming. The energy-expenditure is potentially a measurable quantity. It can also apparently enable, as a common ground, the translation between temporal and spatial models.

In multimodal meaning-making, the mutual support between modalities is the default condition. Under this regime, the information expressed by different modalities is semantically coherent; the modalities are therefore *close* in the terms of the spatial model. In the occasion of multimodal conflict, additional processing effort is required. The need for energy abruptly rises, in order to overcome the semantic *distance*. On the basis of this model, two types of semiotic behavior can be distinguished. The first is driven by the goal of optimizing energy expenditure; the second is driven by the goal of processing new information. The second is relatively energy-consuming, because it poses a requirement for an active synthesis.

Considering the variations in semiosis and the orientation towards the creation of new meanings, let us finally touch upon some examples of modality conflicts in film, discussing the topic of incongruent music in recent film theory and analyze briefly how internal diegetic sound (IDS) is employed in Laurence Olivier's *Hamlet* (1948).

3.5 Functions of audiovisual conflict: from incongruent music to internal diegetic sound

Music and other sounds have accompanied visual content from cinema's earliest stages of development, bringing along the problems of structural and semantic congruence between auditory and visual information (see, e.g., Altman 2004; Ireland

2018). In film's "silent" period, modality conflicts could easily emerge just from the nature of unsynchronized and unrecorded sound, and the inconsistent character of live musical accompaniment.

The majority of the works discussing modality conflict comes from the field of film music research, orienting itself towards the relationship between film and its soundtrack. A seminal approach by Claudia Gorbman notably distinguishes between three basic ways that music can "mean" or signify, namely, via *cultural*, *musical*, and *cinematic* codes (1987: 2, 3). While Gorbman focuses on narrative films, in which the main function of music is hiding films' materiality or constructedness (1987: 58), since Marshall and Cohen's (1988) influential study of how soundtracks guide the perception of short animated films, *congruence* and, respectively, *incongruence* have been established as relevant terms when discussing the psychological influence of music in multimedia (Ireland 2018: 30, see also Willemsen and Kiss 2013). It should be noted that Michel Chion's term *anempathetic music* (Chion 2021: 244) addresses a similar sphere of problems.

Notably, several important studies approach their object rather broadly. Kay Dickinson's *film-music mismatch* verges towards the metaphorical, as she discusses film-making and viewership from the standpoint of sociology, Hegelian dialectics, and cultural production (Dickinson 2008: 31, 34). When Dickinson's *mismatch* points to the cases where film-music relationship "doesn't work" according to aesthetic or ethic sociocultural prerequisites (Dickinson 2008: 14), she makes profound notes about the social context of cinema's development but leaves the closer details of the *mismatch* haunted by a certain subjectivity. In a similar manner, David Ireland's effort to redefine incongruity as the lack of shared properties in auditory and visual modality (Ireland 2018: 34), in his own words, "does not stipulate on which dimensions of the audiovisual relationship the lack of shared properties may be identified: therefore, it facilitates analysis of holistic, subjective judgements of (in)congruity and localised audiovisual difference that may influence these judgements" (Ireland 2018: 34).

Thus, in my view, it would be beneficial if we consider an alternative focus for a change. On the one hand, discussing auditory and visual modalities instead of a film-music dichotomy[16] may seem even more general, but on the other, accepting the auditory modality as consisting of music, sound, and speech as modes with their medial variants and submodes (see Stöckl 2004: 13) allows us some more precision, while such a multimodal approach can remain perfectly complementary with musicology-inspired perspectives. In addition to music, noises, sound effects, and

16 *Film versus music* duality has another subtle issue. As a logical construction, it somewhat transgresses the borders of categories: film tends to entail music; music is not in a syntactic relationship with film as a whole, but is a part of it.

characters' speech can be important sources of incongruity. When modelling a meaning-making process in a complicated, multi-component text, acknowledging that conflicts emerge between modalities and modes instead of just film and music, offers us a more detailed framework.

Studies of incongruent film music have spotlighted various purposes and functions that such a juxtaposition serves. Both Ireland (2018: 94) and Willemsen and Kiss (2013: 171) point out a frequent practice of ironic comment, notably in Stanley Kubrick's films, especially in *A Clockwork Orange* (1971), in which Malcolm McDowell's character hums "Singing in the rain" while torturing his victims. In Kubrick's own words, "It was necessary to find a way of stylizing the violence, just as Burgess does by his writing style. The ironic counterpoint of the music was certainly one of the ways of achieving this" (Kubrick via Nelson 1982: 134).

Among other purposes, Ireland's overview refers to a connection between defamiliarizing effect and reflexivity of an artwork (Ireland 2018: 87, 88): the conflict disrupts an illusion of the seamless narrative world, referring to the constructedness of the text. It is crucial to notice that such effects are often more complex than just an incongruent relation between music and image; the meaning of a film segment is brought forward via various details and agencies, including the work of actors, camerapersons, screenwriters, editors, sound editors, and many more. The semiotic resources that are being manipulated in this process can be seen as modes that, in turn, are organized by visual and auditory modalities. Perhaps one of the most notable examples of self-reflexive defamiliarization in film history relies upon a character's speech. In Jean-Luc Godard's *Pierrot le Fou* (1965), Jean-Paul Belmondo's character suddenly turns towards the camera and addresses the viewer. And not only so: when Anna Karina's character asks whom he talked to, he answers, "*spectateur*."

Last but not least, as a modest counterbalance for music-oriented studies, let us briefly refer to a device called *internal diegetic sound* (IDS). Instead of music, it employs character's speech, but still relies upon the conflict between visual and auditory modalities. IDS happens when the physcial source of the speech is in the film scene, but the character is not visibly speaking; the sound "comes from inside the mind of the character" (Bordwell et al. 2019: 291). In most cases, IDS represents character's inner speech; it is heard only "in the head" of a character and, of course, by the viewer. As such, it conveys character's thoughts as they happen in the time and place of the diegesis (Horton 2017: 194, 195). As a device of meaning-making, it helps to convey personally experienced, subjective, Merleau-Ponti-esque space (Huvenne 2017: 51).

Briefly mentioned by Bordwell and his colleagues (Bordwell et al. 2019: 291) as an example of IDS, Hamlet's monologue in Laurence Olivier's film[17] features a transfer

17 Olivier directed the film and also starred as Hamlet.

between IDS and "normal" speaking. During the first seconds of the monologue, the camera rests on the roaring sea waves that dissolve to the eyes of Olivier. After this, a sudden cut to medium close-up of him sitting on a rock, both seen and heard speaking. Up to that moment, visual and auditory modality have supported each other. With the lines "… take arms against a sea of troubles … and by opposing, end them," Olivier draws a dagger from his belt, holding it at the height of his chest. Some quiet and sombre music starts. Just then, the most interesting thing happens. At approximately 01:03:10, Olivier closes his eyes, and with his mouth also closed, the auditory modality continues to carry the monologue. Consequently, the words "To die. To sleep no more" become the first lines of IDS, although there is already some internality to a man speaking loudly to himself in a lonely place.

The camera dollies in, framing Olivier's troubled face, eyes closed, sweat-drops on his forehead revealed as highlights. A similar close-up is presented as before, namely, the actor's eyes, and the soundtrack goes on "…it is a consummation devoutly to be wished. To die. To sleep. To sleep." At 01:03:35, the music suddenly bursts into a violent quaver of string instruments. This may both eerily and comically suggest an alarm clock. After a bout of intensive music, Hamlet "wakes up," opens his eyes, stirs his body and resumes to declamate his monologue in both auditory and visual modality. The music subsides and stops, leaving only the murmur of the sea in the background. At 01:05:05, as the sequence nears its end, Hamlet drops the dagger down the cliff. The film is cut to the next shot, deep downward angle towards the waves, dagger falling into the abyss. Then, a cut back to Hamlet, who delivers his final lines, looking briefly down the cliff and walking away from the camera. The music silently starts again, image fades to black, concluding the scene.

As a possible interpretation, I suggest that the staging of the monologue can be seen in the Jungian paradigm of hero's journey, reflecting the stages of *the refusal of the call* and the *crossing of the first threshold*, widely followed by many storytellers (see, e.g., Campbell 1993 [1949]: 59, 77). In this case, the disruption of modalities marks a *critical turning point* in the hero's progress. Initial hesitation and instability is followed by a dreamlike episode where the old perspective is transformed into a new, decisive, and active one. It should be kept in mind that in the next segment, Polonius announces the arrival of the actors, giving Hamlet the ideas and "tools" for the revenge. During the liminal, transformative stage, the semantic discontinuity between auditory and visual modalities correlates with Hamlet's hesitation. When the distance has been finally overcome, the hero has fresh objectives, and the narrative has moved into the next phase. In this case, IDS functions mostly as device of punctuation and accentuation, marking a turning point in the story.

4 Concluding thoughts

The example of Hamlet's monologue (as well as those of J-cuts and L-cuts that contribute to continuity editing instead of disrupting it) refer to an important point: technical or formal modality conflicts don't have to necessarily function as semantic conflicts. In the case of internal diegetic sound (IDS), perceiving a character's closed mouth and hearing his voice at the same time, might be startling only to a viewer with little cultural experience. If a device works smoothly (as IDS does for representing characters' thoughts or inner speech), the repeated use of it diminishes artistic effect and decreases novelty; it is quickly automated and becomes a part of the specific sign system that has been metaphorically referred to as "film language."[18] In other words, it becomes culturally coded and is deprived of its status as a fluid, volatile "third meaning." Similarly, discussing the rather broad field of incongruent film music, we may encounter instances where technically incongruent-sounding music fails to provoke a semantic conflict.

The present article's status as an introductory observation of modality conflicts inclined the main focus towards the formal description of the conflict. Thus, the next logical step in addressing the topic would be a further elaboration of the semantic dimension, exploring more thoroughly the capacity of description that is offered by spatial models, not only Lotman's, but also, for instance, those that are utilized for semantic modeling in computer sciences. Developing a more comprehensive understanding of the relation between formal and semantic incongruences would be essential.

The notion of *dominant* discussed by Jakobson should be seen as equally pertinent. In dynamic (sign) systems, different components (as Lotman's languages, modes as semiotic resources or sensory modalities in multimodal texts) occupy the dominant position in turn, taking the role as the sources of the organizing principles for the meaning-making. Simultaneously, the concept of dominant requires careful and critical consideration in the light of the fact that in the synthetic meaning-making and the emergence of new structures, the old ones could undergo a significant transformation and surrender their properties, including the dominance over the system.

Multimodal conflicts (especially semantic ones) create the situation where two or more possibilities for meaning are juxtaposed. A lot of theoretical approaches towards such situations have their roots in Hegel's dialectics, specifically moving along the chain *affirmation – negation – negation of negation,* which Marx later retitled *thesis – antithesis – synthesis* (see, e.g., Dickinson 2008: 31). Both Eisenstein's

18 I still prefer to use these notions in parentheses, because the structuralist, Metzian parallels between the filmic semiosis and natural language can be deemed problematic.

concept of *intellectual montage* and Lotman's idea of *incompatible languages* as the mechanism for creating new information can be seen as the developments of such dialectics. For Lotman's mechanism, the dynamism is an especially important characteristic. When two incompatible messages are juxtaposed, the perceiver of the text is activated; she has to choose between variants or synthesize an emergent meaning, thus participating more actively in the semiosis. For Lotman, as was for Eisenstein, this is the core artistic mechanism which is also exemplified by the Russian Formalists' concept of *ostranenie* or defamiliarization.

In film, the mutual interaction of image and sound can be seen as the bimodal synthesis of visual and auditory components and discussed in the framework of multimodal film semiotics. For evolutionary reasons, the supportive relation between image and sound is regular practice of meaning-making. The modality conflict, especially when its semantic potential is fulfilled, is rather a special case. As a semiotic strategy, it can be employed with various purposes in mind, from ironical comments to the fundamental prompts for completely alternating the viewing regime. In cognitive approaches, the concept of *bodily simulation* has a central position. The multimodal conflict can work against simulation; it has the potential to encourage the viewer to generate her subjective meanings.

It has to be concluded that the notion of *conflict* does not refer to something simple and self-evident. While Eisenstein talked about the *juxtaposition*, the potential for conflict still greatly varies in such synthetical practices. For that reason, the article discussed the possibility of measuring the scope of alleged modality conflicts. First the concepts of *redundancy* and *synchrony* were touched upon, with the conclusion that neither can be considered as a basis for such measuring. Following Altman and Chion, we can recognize that auditory and visual modalities are not mutually redundant. Therefore we cannot define modality conflict by the disruption of a "normal" state of redundancy. The similar problem is with synchrony: paradoxically, *synchrony is usually required* as the basis for modality conflicts.

As suggested above, Lotman's idea of *space* as a complementary primary modeling system would be helpful in evaluating the scope of multimodal conflicts. As spatial models feature *distance* as the main parameter, in modality conflicts the space should be approached metaphorically: the spatial distance is set to represent *semantic distance*. Still, it is only a half-way journey, because *semantic distance* brings along its own problems, e.g., entanglement in the contextual influences and the subjective interpretations by viewers.

Nevertheless, the spatial model allows to consider a mechanism that underlies the translation between the dimensions of *time and space*, and *auditory and visual*. The parallels between conceptual, semantic space, and space in the brain architecture can be noticed. Both are characterized by the concern for energy expenditure. This, in turn, can by hypothetically seen as the common feature that unites different

aspects of meaning-making: using the temporal and spatial models, and also operations that involve the work of the memory. I suggest those ideas are worthy of further discussion and empirical approaches.

References

Altman, Rick. 1980. Moving lips: Cinema as ventriloquism. *Yale French Studies* 60. 67–79.

Altman, Rick. 2004. *Silent film sound*. New York: Columbia University Press.

Anderson, Joseph D. 1996. *The reality of illusion: An ecological approach to film theory*. Carbondale & Edwardsville, IL: Southern Illinois University Press.

Barthes, Roland. 1977. The third meaning. In *Image music text*, 52–68. London: Fontana Press.

Bateman, John & Karl-Heinrich Schmidt. 2011. *Multimodal film analysis: How films mean*. New York & London: Routledge.

Bateman, John, Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality: Foundations, research and analysis. A problem-oriented introduction*. Berlin & New York: Mouton De Gruyter.

Bennett, Tyler James. 2021. Incompatibility, unlimited semiosis, aesthetic function. In Elin Sütiste, Remo Gramigna, Jonathan Griffin & Silvi Salupere (eds.), *(Re)considering Roman Jakobson* (Tartu Semiotics Library 23), 149–163. Tartu: University of Tartu Press.

Benveniste, Émile. 1981. The semiology of language. *Semiotica* 37. 5–23.

Bertelson, Paul & Béatrice De Gelder. 2004. The psychology of multimodal perception. In Charles Spence & Jon Driver (eds.), *Crossmodal space and crossmodal attention*, 141–178. Oxford: Oxford University Press.

Bordwell, David, Kristin Thompson & Jeff Smith. 2019. *Film art: An introduction*, 12th edn. New York: McGraw-Hill.

Boria, Monica & Marcus Tomalin. 2020. Introduction. In Monica Boria, Ángeles Carreres, María Noriega-Sánchez & Marcus Tomalin (eds.), *Translation and multimodality: Beyond words*, 1–23. London & New York: Routlegde.

Campbell, Joseph. 1993 [1949]. *The hero with a thousand faces*. London: Fontana.

Chion, Michel. 2016. *Sound: An acoulogical treatise*. Durham & London: Duke University Press.

Chion, Michel. 2021. *Music in cinema*. New York: Columbia University Press.

Dickinson, Kay. 2008. *When film and music won't work together*. New York: Oxford University Press.

Eisenstein, Sergei. 1957. *The film sense*. New York: Meridian.

Eisenstein, Sergei. 1961. One path to color. *Sight & Sound* 30(2). 84.

Eisenstein, Sergei. 1988 [1923]. The montage of attractions. In Richard Taylor (ed. & trans.), *Writings 1922–1934*, vol. 1, 33–38. London: BFI.

Eisenstein, Sergei, Vsevolod Pudovkin & Grigori Alexandrov. 1977 [1928]. Statement on sound. In Jay Leyda (ed. & trans.), *Film form: Essays in film theory*, 257. New York & London: HJB.

Elleström, Lars. 2020. The modalities of media II: An expanded model for understanding intermedial relations. In Lars Elleström (ed.), *Intermedial relations among multimodal media* (Beyond media borders 1), 3–91. Cham: Palgrave MacMillan.

Forceville, Charles. 2020. *Visual and multimodal communication*. New York: Oxford University Press.

Frierson, Michael. 2018. *Film and video editing theory: How editing creates meaning*. New York: Routledge.

Gallese, Vittorio & Michele Guerra. 2015. *The empathic screen: Cinema and neuroscience*. Oxford: Oxford University Press.

Gibson, James J. 2015 [1979]. *The ecological approach to visual perception*. New York: Psychology Press.

Gleick, James. 2011. *The information: A history*. London: Fourth Estate.
Goodwin, James. 1993. *Eisenstein, cinema, history*. Chicago, IL: University of Illinois Press.
Gorbman, Claudia. 1987. *Unheard melodies*. Bloomington, IN: Indiana University Press.
Grodal, Torben. 2009. *Embodied visions: Evolution, emotion, culture, and film*. Oxford: Oxford University Press.
Hayward, Susan. 2000. *Cinema studies: The key concepts*. London & New York: Routledge.
Horton, Justin. 2017. Sound, space, and complex narrative cinema. In Marta Boni (ed.), *World building: Transmedia, fans, industries*, 187–203. Amsterdam: Amsterdam University Press.
Huvenne, Martine. 2017. Editing space as an audio-visual composition. In Janina Wildfeuer & John Bateman (eds.), *Film text analysis: New perspectives on the analysis of filmic meaning*, 46–65. London & New York: Routlegde.
Ireland, David. 2018. *Identifying and interpreting incongruent film music*. Cham: Palgrave Macmillan.
Jakobson, Roman. 1971a [1963]. Visual and auditory signs. In *Selected writings II: Word and language*, 334–337. The Hague & Paris: Mouton.
Jakobson, Roman. 1971b [1964]. On the relation between visual and auditory signs. In *Selected writings II: Word and language*, 338–344. The Hague & Paris: Mouton.
Jakobson, Roman. 1971c [1967]. Language in relation to other communication systems. In *Selected writings II: Word and language*, 697–708. The Hague & Paris: Mouton.
Jakobson, Roman. 1981 [1935]. The dominant. In *Selected writings III: Poetry of grammar and grammar of poetry*, 751–756. The Hague: Mouton.
Jullier, Laurent. 2010. Should I see what I believe? Audiovisual ostranenie and evolutionary-cognitive film theory. In Annie van den Oever (ed.), *Ostrannenie: On "strangeness" and the moving image. The history, reception, and relevance of a concept*, 119–140. Amsterdam: Amsterdam University Press.
Kress, Gunther. 2010. *Multimodality: A social semiotic approach to contemporary communication*. Milton Park: Taylor & Francis.
Kress, Gunther. 2020. Transposing meaning: Translation in a multimodal semiotic landscape. In Monica Boria, Ángeles Carreres, María Noriega-Sánchez & Marcus Tomalin (eds.), *Translation and multimodality: Beyond words*, 24–48. London & New York: Routlegde.
Kress, Gunther, Carey Jewitt, Jon Ogborn & Constantinos Tsatsarelis. 2000. *Multimodal teaching and learning*. London: Continuum.
Kull, Kalevi. 2015. Semiosis stems from logical incompatibility in organic nature: Why biophysics does not see meaning, while biosemiotics does. *Progress in Biophysics and Molecular Biology* 119(3). 616–621.
Lotman, Jurij. 1977. *The structure of the artistic text*. Ann Arbor, MI: University of Michigan Press.
Lotman, Yuri. 1990a. Artistic space in Gogol's prose. *Russian Literature Triquarterly* 23. 199–241.
Lotman, Yuri M. 1990b. *Universe of mind*. London: L. B. Tauris.
Lotman, Yuri. 2009 [1992]. *Culture and explosion*. Berlin & New York: Mouton De Gruyter.
Lotman, Juri, Vyacheslav Ivanov, Aleksandr Pjatigorskij, Vladimir Toporov & Uspenskij Boris. 2013 [1973]. Theses on the semiotic study of cultures (as applied to the Slavic texts). In Silvi Salupere, Peeter Torop & Kalevi Kull (eds.), *Beginnings of the semiotics of culture* (Tartu Semiotics Library 13), 53–77. Tartu: University of Tartu Press.
Marks, Laura U. 2000. *The skin of the film: Intercultural cinema, embodiment, and the senses*. Durham & London: Duke University Press.
Marshall, Sandra K. & Annabel J. Cohen. 1988. Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception* 6(1). 95–112.
Min Hong, Seung. 2019. Contrapuntal aurality: Exceptional sound in Hollywood monster horror films during the early sound era. *Journal of Popular Film and Television* 47(4). 215–226.
Nelson, Thomas A. 1982. *Kubrick, inside a film artist's maze*. Bloomington, IN: Indiana University Press.

Oja, Martin. 2014. Darkness on screen: Subjectivity-inducing mechanisms in contemporary Estonian art film. *Baltic Screen Media Review* 2. 76–95.

Richards, Rashna Wadia. 2008. Unsynched: The contrapuntal sounds of Luis Buñuel's *L'Age d'or*. *Film Criticism* 33(2). 23–43.

Robertson, Robert. 2009. *Eisenstein on the audiovisual: The montage of music, image, and sound in cinema*. London: Tauris.

Sebeok, Thomas. 1991. In what sense is language a "primary modeling system." In Myrdene Anderson & Floyd Merrell (eds.), *On semiotic modeling*, 327–229. New York & Berlin: Mouton de Gruyter.

Shams, Ladan & Robyn Kim. 2010. Crossmodal influences on visual perception. *Physics of Life Reviews* 7. 269–284.

Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423.

Smith, Tim J. 2013. Watching you watch movies: Using eye tracking to inform film theory. In Arthur Shimamura (ed.), *Psychocinematics: Exploring cognition at the movies*, 165–191. New York: Oxford University Press.

Sobchack, Vivian. 2004. What my fingers knew: The cinesthetic subject, or vision in the flesh. In *Carnal thoughts: Embodiment and moving image culture*, 53–84. Berkeley, CA: University of California Press.

Stam, Robert, Robert Burgoyne & Sandy Flitterman-Lewis. 1992. *New vocabularies in film semiotics: Structuralism, post-structuralism, and beyond*. London: Routledge.

Stöckl, Hartmut. 2004. In between modes: Language and image in printed media. In Eija Ventola, Charles Cassily & Martin Kaltenbache (eds.), *Perspectives on multimodality*, 9–30. Amsterdam & Philadelphia: John Benjamins.

Taberham, Paul. 2013. Correspondences in cinema: Synaesthetic film reconsidered. *Animation Journal* 21. 47–68.

Torop, Peeter. 2022. Lotman's semiotics of literature in terms of "space as language". *Neohelicon* 49. 581–591.

Willemsen, Steven & Miklós Kiss. 2013. Unsettling melodies: A cognitive approach to incongruent film music. *Acta Universitatis Sapientiae, Film and Media Studies* 7. 169–183.