

DARIO MARTINELLI

## FILMING THE NOTES: AUDIOVISUAL STRATEGIES IN POPULAR MUSIC COMPOSITION AND PRODUCTION

### INTRODUCTION

The relationship between audiovisuality and popular music has been addressed in many ways, and according to many approaches, and it is definitely not the intention of this essay to revisit—critically or otherwise—these contributions. I will merely mention the three groups (plus subgroups) of approaches that I consider most common.

Certainly, the most typical approach focuses on when, where and how popular music was used in audiovisuality (on popular music and advertising see, e.g., Klein 2009). Privileged subcategories of this group are those concerning audiovisual texts specifically designed to accompany, support or promote popular songs or repertoire, such as musicals, video clips, rockumentaries and others (e.g., Frith et al. 1993). Next, there have been numerous discussions about the specific relationship between directors and popular music, or between popular musicians and audiovisual texts, or even between composers of soundtracks and popular music (as I myself did in, e.g., Martinelli 2008 and 2014). Finally, scholars have extensively talked about the audiovisual representation of popular music as a (social, cultural, etc.) “phenomenon” (e.g., James 2016).

Significantly fewer reflections have been produced about the presence in popular music of operative and stylistic strategies that are more typical, if not actually defining, of audiovisuality. Not that the topic is ignored—rather, it is taken for granted, even by the authors themselves, that a song, also an

---

Prof. dr. DARIO MARTINELLI, Kaunas University of Technology, Lithuania; e-mail: [dario.martinelli@ktu.lt](mailto:dario.martinelli@ktu.lt); ORCID: <https://orcid.org/0000-0002-2297-2308>.

intrinsically narrative art form, can share a series of representation strategies with an audiovisual text. When discussing his track “Band on the Run”, for example, Paul McCartney has often described the compositional approach of the piece (having a clear tripartite structure, like a suite), as that of a film in three sequences: the imprisonment of the protagonists (the part starting with “Stuck inside these four walls...”), the escape plan (“If I ever get out of here...”) and finally the actual escape (“Well, the rain exploded with a mighty crash as we fell into the sun...”). Some musicologists can naturally elaborate a little on the concept, as for example we see in Fabbri (2008, 203–6) with the Italian song “Se telefonando”.

The problem is that these references are always quite generic and usually focused on wide-ranging filmological issues: narration, editing, general structure.... When and if one tries to go into the merits of given strategies of audiovisual representation, we begin to feel the lack of adequate and exhaustive literature.

#### LUCA MARCONI'S IMPORTANT CONTRIBUTION

Luca Marconi, whom we sadly lost in 2019, was an active member of the Musical Signification Project, and an esteemed colleague and good friend of mine. In the auspicious occasion of the MSP's 40th anniversary, I am particularly happy to pay a small tribute to him by adopting one of his theoretical reflections (2009) as my first case study.

Marconi's study is based on the thesis that “each song is a syncretic text whose relationship with the system of genres and styles it belongs to is manifested through the invitation, addressed to a *model listener*, to share the values that guide the actions on which it focuses its attention” (2009, 85; my translation). The main aim of his essay is to develop “a typology of the subjects of the actions the songs focuses on” (85), on both a purely musical and textual level (including the area of interaction of the two levels). Marconi classifies these subjects according to five categories (85–94), applied to the Italian song “Il cielo in una stanza” (I will provide here examples taken from a more international repertoire). The categories are:

1) *Songs focused on the speaker*. The speaker is not the singer/performer in the strict sense. It is rather the “speaking” (and singing) voice of a song, which tells us things because the song's melody and lyrics require that someone tells us. The speaker of a song written and sung by a singer-songwriter,

thus, is not necessarily the latter, but can be anyone. Browsing through the list of traits pertinent to this type of focus (86–88), a particularly telling example is “the speaker’s frequent use of interpellations, which simulate a direct communication that resembles a call” (87). In popular music we very often find such forms of direct communication (interjections, appellatives, endearments) inserted in melodic contexts that effectively simulate their use in speech: in Crowded House’s “Don’t Dream It’s Over”, for instance, we find the passage “Hey now, hey now, don’t dream it’s over”, where the double “Hey now” is conveyed through two different melodic figures, the second being higher in pitch, that well mimic the prosody of a spoken encouragement (where usually, the repetition of words like “come on” or indeed “hey now”, is in a sort of crescendo, as if to say “you can do it, you can *really* do it!”).

2) *Songs focused on the protagonist*. Whether the author speaks in the first person and therefore appears as the (autobiographical or fictitious) song’s protagonist, and—above all—if the song is in the third person, possibly citing names, places and times, the listeners perceive a story, with characters, events, and can be directed to focus on those. Among the countless examples, we could perhaps pick Pink Floyd’s “The Gnome”, which also features a story-telling structure in the lyrics (“I want to tell you a story, about a little man, if I can...” ) and then proceeds to name (“A gnome named Grumble Grumble...” ) and describe the protagonist (“He wore a scarlet tunic, a blue green hood...” ).

3) *Songs focused on the coryphaeus*. The coryphaeus was the leader of the choir in the Greek theater. With this charmingly archaic definition (one could also go with “conductor”, in principle), Marconi refers to songs that exercise a conative function on the listener, inviting them to react, especially in a senso-motorial sense. For instance, every song inviting us to move the body or dance tends to activate this typology. The singing voice often gives instructions to the listener, indeed “conducts”, on how/when/where/with whom to dance. For instance, in Chubby Checker’s famous “Let’s Twist”, we hear “Come on everybody, clap your hands... We’re gonna do the twist and it goes like this... Come on let’s twist again...”.

4) *Songs focused on the author*. Borrowing from Merleau-Ponty’s (1960) phenomenology, Marconi specifies that in the previous three categories we have dealt with “transparent” semiotic instances, that is, on a focusing of the actions signified by the songs themselves (trivially: what a song talks about). In this fourth category, the instances become “opaque”, that is, the attention now focuses “on the productive actions that can be inferred by considering their characteristics” (Marconi 2009, 91). When listening to a song, we want

to know how autobiographical, authentic, commercial, committed etc. it is, in the author's intentions. For example, when listening to a song like Tori Amos's "Me and a Gun", which Amos herself admitted being about the sexual abuse she fell victim to in her youth, we get a direct (or rather poetry-mediated) access to the author's feelings and experience.

5) *Songs focused on the performer*. This typology includes all the versions of a given song that appear after a chronologically first, conventionally "original", version. The practice of publishing *other* versions of a song has been constantly increasing in variety. To the traditional types (live performances and covers) items like remixes, remakes, remasters, mashups and others have been added. In such situations, meanings and associations can be attributed *starting from* the performer: it is the performance itself that dictates the mood and the semiotic clues to interpret the piece. "(You Make Me Feel Like) A Natural Woman", a song that is well known in at least two versions, was written by the legendary Goffin-King duo, and Carole King herself made her own version in the album *Tapestry*, in 1970. Three years earlier, however, the most famous and, for many, ultimate version of the song was released by Aretha Franklin. A comparison between the two versions will immediately display what Marconi means. Franklin, one of the most powerful voices in the business, made an outspoken, passionate version of the song, encouraging the listener to interpret the lyrics as the celebration of a woman's bodily (and possibly social, too, Franklin being an Afro-American singer performing the song in full Civil Rights movement times) liberation. Franklin's "natural woman" is a woman who can freely express herself, her exuberance, her identity, her (metaphorical and factual) nudity to anybody. A rich arrangement, in classic Atlantic Records style, with a full band and a horn section, contributes to this, too.

When we instead take King's version, we are exposed to a piano-and-voice-only arrangement, a modest, almost shy singing performance from a white (and, we might add, "middle-class" in spirit) singer. Suddenly, the "natural woman" is now a woman dealing with her own insecurities and fears, who, thanks to the man addressed in the song (the "you" of the lyrics), has the chance to share her personality and does not need to hide them. The song, in other words, has achieved a very intimate profile, the introvert Yin to the extrovert Yang of Franklin's version.

## PRODUCTION AND MONTAGE

Beyond the intrinsic merits of Marconi's analysis, what seems particularly noteworthy is that the five focuses can be easily translated into the audiovisual field. Indeed, in some cases they seem *particularly* pertinent to audiovisuality. In my 2023 paper, for instance, I had the opportunity to discuss how the five characters of Marconi's categories move within and without the conceptual space of a song, just like directors, actors and the characters the latter play do in a film. In other words, we witness the existence of a diegetic and non-diegetic space in songs as well.

Moreover—and such will instead be my second case study here—the necessity, in a song's arrangement and production, to direct the focus to this or that character, combined with several other narrative and semantic needs that a song possesses, raises also questions within a field that has been considered intrinsic and defining of audiovisuality as a whole. I am of course talking about “montage”.

As Ingmar Bergman once said, “montage is akin to music: it is the playing of the emotions”. We generally define it as the act of editing, cutting, and piecing together a text in such a way that its communicative potential is the most faithful possible to what the authors intended to express, at both an explicit and a more metaphorical-emotional level. Montage employs images and sounds as raw material, and puts them together in different ways, often violating spatial and temporal coordinates (for instance by placing two chronologically distant images one after another, as if happening in sequence). Montage is the quintessential filmic action, whose ultimate asset lies in its ability to transform a text into a dense network where just about every single item can establish a relationship with another.

Music production, in this sense, is really filmic montage minus images. What is particularly relevant is that the great added value of both production and montage lies in its ability to enhance the communicative potential of a text. That occurs because both practices are not simply a sum of parts but are actually able to create new meanings out of such sums. If a producer aligns a verse like “I heard you on the wireless back in ‘52” (of course, I mean the 1979 smash hit “Video Killed the Radio Star”) with a lo-fi voice centered on the mid frequencies, as if indeed coming from an old, portable radio, the result is not only a mere sum of a verse and a sound, but also a relation that is established/encouraged through this association: a young boy listening to the radio in the 1950s, probably in his room, with the sound

quality of a cheap radio set (thus, we also get information on the social status of the boy).

The Latvian-Soviet great director Sergei Eisenstein, who was also a prominent film theorist, compared montage to the compounding of characters in Japanese writing. When one combines the character for “dog” to the character for “mouth”, he noted, the result is not only “dog’s mouth” but also the new concept of “bark”: similarly, the nature of production is “dialectical”.

Generally speaking, one can distinguish among three “pure” types of montage: narrative, graphic, and ideational. I say “pure” in quotation marks because in actual fact, these types rarely appear alone, but they are more likely combined in different fashions. In the narrative type, montage is at the explicit service of the story: the idea is to follow a subject/event from one point to another, by offering multiple points of views and angles, that allow a better understanding of that portion of the story, add more information, and often prevent the sequence from being too repetitive and monotonous. In production, we may understand “narration” in two senses: a conventional lyrical-thematic sense, where an actual story, expressed linguistically, takes place; and a more abstract structural sense, related to the musical development of a piece. In the former sense, the production practically follows the words, developing strategies to represent the song’s meanings. The catalogue of ideas and sounds on Supertramp’s “Logical Song”, for instance, with the various rhythmic patterns played by the instruments in a kind of *Modern Times* clockwork/production chain blend, or the whistle heard to introduce the coda, are all intended to contribute to this description of “logical”, hyper-rational, schematic life that the song’s protagonist feels alienated by. In the latter sense, the production “follows the structure”, that is, enacts the intrinsically musical narrative that the song format itself dictates. For example, introducing an instrumental and/or vocal break before an upbeat/energetic part is one of such strategies, because it metaphorically corresponds to “catching a breath” or “taking a run up” before a race or a jump. We often see that in The Beach Boys, with their harmonized vocal breaks without instruments (e.g., the “Let’s go surfin’ now, everybody’s learnin’ how, come on and safari with me” part in “Surfin’ Safari”).

In the graphic type, shots are not juxtaposed on narrative grounds but aesthetic ones. Associations can be made by color, sound, shape, or anything that has more a formal value than a specific role in the story. Still, the latter exists anyway, because those associations still contribute to (in fact, enrich) the overall meaning and significance of the text. When it comes to production, we can probably say that the “graphic type” is the most typical strategy, or at least

the one presenting most episodes within a single production. Indeed, in its basic assumption, production remains a process of “putting sounds together” in order to produce an aesthetically satisfying result. Criteria that contribute to this process may include the available studio and performers’ resources, the personal taste, the genres/styles of reference, some historical/technological variables, and some geographical/cultural influences. We could take the case of the piccolo trumpet solo on The Beatles’ “Penny Lane”. Here, first and foremost, we have an aesthetic choice: a piccolo trumpet playing a Bach-inspired solo has nothing intrinsically (narratively) to do with the Liverpool district Paul McCartney is referring to in his song. What comes into play, instead, is at least the following: a number of available resources (enough budget to hire a guest musician, McCartney’s instinctive sense of melody, Martin’s ability to transcribe that melody, etc.), personal taste (McCartney had developed the idea after listening to Bach’s *Brandenburg Concerto* no. 2), genre references (the song operates within the realm of so-called Baroque rock, that The Beatles themselves had contributed to initiate), and obviously geographical/cultural aspects (as the song flirts with European art music, taking therefore a more intellectual edge).

Finally, in the ideational type, montage serves that purpose so cleverly described by Eisenstein with his comparison to Japanese characters: the juxtaposition/sum of two (or more) elements generates a new meaning, an idea, a metaphor, an atmosphere, etc. We have seen that already with the additional information provided by the lo-fi voice in “Video Killed the Radio Star”.

#### ARRANGED AND PRODUCED BY... SERGEY EISENSTEIN

This general tripartition has been deepened and detailed in various ways over the years by film scholars, but the five-point classification developed by Eisenstein in works such as 1942 and 1949 (I refer here to the English translations) remain one of the most accurate. Eisenstein’s theory of montage is “dialectical”: it is characterized by confrontation and conflict between the items involved. Montage can be seen as the creation of a theme from the juxtaposition of specific details. The audience needs also to employ its own “encyclopedia” (as a complex system of shared knowledge that governs the production and interpretation of signs inside communicative contexts—see Eco 1979) to interpret the result of such confrontation, and interface it with

the “suggestion” of the author. In this way, the author and the spectator/listener cooperate in creating the text’s narrative and aesthetic structure.

1. *Metric montage*. This typology is characterized by the juxtaposition of film bits in which the main criterion for construction is the actual length of the various bits. Each bit is joined together relying purely on the physical nature of time, cutting to the next bit no matter what is happening within the image (or the sound). The length of the bits can be also mathematically calculated according to a metric formula: the lengths as such may vary, but they maintain the original proportions of the formula, obtaining (depending on the case, and therefore on the formula) a wide range of emotions, from serenity to tension, from a slow to a frenetic pace. A metric approach may not be the most common strategy in production, but we find an excellent example in The Beatles’ “Tomorrow Never Knows”, namely the employment of the various loops that characterize the track. Each loop was inserted in the recording in accordance with its own length, with no specific connection to any given point of the song, and with no time manipulation that would slow it down or speed it up.

2. *Rhythmic montage*. Here, the length of the bit derives from the specifics of the bit itself and the sequence in general. Tension can be achieved by the confrontation between the length of the bit and the movement within it. A rhythmic type of production is a very common strategy, since songs display an immanent narrative structure through their form. Instruments, for example, may appear and disappear depending on the length of each part. Typically, in a strophe-refrain structure, the instrumentation is differentiated in the two parts, usually with the purpose of making the refrain sound more poignant: for instance, in Bob Dylan’s “Knockin’ on Heaven’s Door”, the backing vocals turn from a gospel-like “Ooh” in the background during the strophes into harmonizing the words “Knock-knock-knockin’ on heaven’s door” in the refrain, and the drums turn from playing the upbeat accents on the snare rim into playing them on the center of the snare—both solutions giving more power to the refrain.

3. *Tonal montage*. This typology uses the emotional meaning of the various bits. In this case, we do not just have a manipulation of their temporal length or rhythmical characteristics, but we aim at eliciting a more complex reaction, and therefore the guidelines, here, are tone and mood, rather than time and tempo. Once again, the above-mentioned arrangements of “Logical Song” and “Video Killed the Radio Star”, namely *the reasons* behind those choices, are good examples of this category.



4. *Overtonal montage*. This typology combines metric, rhythmic, and tonal montages together. Each element derives from each other, aiming at inducing an emotional effect from the audience. Bringing together the various montage methods propels a level of conflict, with each method developing from the other. When the audience is caught into a different range of emotions at the same time that go beyond the “direct” emotion/s created at tonal level, the text will have successfully created an overtonal montage. We obviously face here production at its most complex and articulated, and it is no surprise that we witness instances of overtonal production in those songs that carry multilayered meanings and a high degree of ambiguity at both lyrical and musical level: Queen’s “Bohemian Rhapsody”, Radiohead’s “Paranoid Android”, numerous prog rock songs, etc. An example of in-depth analysis of this case is offered in Martinelli-Bucciarelli (2023, 138–52), as applied to The Beatles’ “The Fool on the Hill”.

5. *Intellectual Montage*. This fifth and final typology of montage aims at moving the spectator psychologically. That is usually achieved by rhetorical associations and contrasts. The emphasis is not upon particular characteristics of the bits as such, but the intellectual process, which they may activate by the way they are assembled, positioned, etc. The process is often predetermined by the director’s personal style and vision, and therefore tends to metaphorically talk about the director themselves as much as it talks about the actual film. This happens in music production as well, where, since Phil Spector inaugurated producers’ personality cult and identifiable individual production styles, we have witnessed numerous producers applying their own artistic imprinting, regardless of the song they were dealing with. A good example is Jeff Lynne, whose signature traits include a descending lick usually performed by an electric guitar at the end of a phrase, several layers of acoustic guitar accompaniments, a particularly thick sound on the drums, and others. The amalgamation of these traits can be considered an “intellectual” process, indeed. Lynne does not think in terms of a particular song as a unique item, with its unique structure, contents, modes of expression. He rather has his own model of production, and a general concept of how a song should sound, and tends to apply them more often than not, making his productions recognizable as his own.

## REFERENCES

- Eco, Umberto. 1979. *Lector in fabula*. Milano: Bompiani.
- Eisenstein, Sergei. 1942. *The Film Sense*. New York: Harcourt.
- Eisenstein, Sergei. 1949. *Film Form: Essays in Film Theory*. New York: Harcourt Brace Jovanovich.
- Fabbri, Franco. 2008. *Il suono in cui viviamo*. Milano: Il Saggiatore.
- Frith, Simon et al., eds. 1993. *Sound and Vision: The Music Video Reader*. London: Routledge.
- James, David. 2016. *Rock'n'Film: Cinema's Dance with Popular Music*. Oxford: Oxford University Press.
- Klein, Bethany. 2009. *As Heard on TV: Popular Music in Advertising*. Farnham: Ashgate Publishing.
- Marconi, Luca. 2009. "Nell'aria e nella stanza: le relazioni tra la musica e le parole nelle canzoni." In *Parole nell'aria. Sincretismi fra musica e altri linguaggi*, edited by Maria Pia Pozzato and Lucio Spaziante, 85–108. Pisa: Edizioni ETS.
- Martinelli, Dario. 2008. "David Lynch, regista suonato." *Moviement* 1:12–21.
- Martinelli, Dario. 2014. "Cinema and Popular Music, a Multifold Relationship. The Cases of Ennio Morricone and David Lynch." *Interdisciplinary Studies in Musicology* 14:155–67.
- Martinelli, Dario. 2023. "A Strophe, a Chorus, and a Bridge Walk into a Bar... : On the Connections between Song Formats and Narrative Strategies." *Chinese Semiotic Studies* 19 (1): 57–77.
- Martinelli, Dario, and Paolo Bucciarelli. 2023. *The Beatles and the Beatlesque: A Crossdisciplinary Analysis of Sound, Production and Stylistic Impact*. Cham: Springer.
- Merleau-Ponty, Maurice. 1960. *Signes*. Paris: Gallimard.

FILMING THE NOTES: AUDIOVISUAL STRATEGIES  
IN POPULAR MUSIC COMPOSITION AND PRODUCTION

Summary

Practices of composing and producing a song bear similarities with filmmaking, particularly the stages of development, *mise-en-scène* and montage. This particular topic, when approached at scholarly level, becomes suddenly superficial. Not that it is scarcely analyzed, but even authors themselves tend to take for granted that a song shares some representation strategies with an audiovisual text. Musicologists can elaborate on these concepts, but their analyses tend to be mere parallels and not rigorous comparisons. When instead we try to elaborate on defined strategies of audiovisual representation, we feel the lack of exhaustive research models.

The present article, along with the author's recent work (particularly Martinelli-Bucciarelli 2023), aims at being an opportunity to (begin to) fill this gap, allowing reflections on specific filmic strategies applied to popular music. The case-studies will be the so-called "focus of the songs" (Marconi 2009) and the application of montage theory (particularly Eisenstein's) on music production.

**Keywords:** production; popular music; audiovisuality; montage