

Daniel Klug

## **What’s That Sound? How Music Videos Create Realistic Audiovisual Experiences**

This paper deals with the way additional sound elements can alter the specific characteristics of the audio-vision of music videos and therefore create types of more “realistic” audiovisual impressions compared to non-mediated audiovisual phenomena.

In this context, “realistic” refers to the perception of everyday audiovisual occurrences that do not involve the use of any communication media and therefore can be classified as non-mediated. However, the audiovisual media artifacts are artificially constructed products in which moving images and sound or music are re-arranged only according to the imagination of a non-mediated counterpart. Compared to other audiovisual media artifacts, for example movies, the audio-vision of music videos is characterized by an inverse combination of music and moving images, which means that moving images are added in orientation to the musical and rhythmical elements of an underlying and pre-existing song.

Depending on the synchronicity of the single visual element and the single acoustic or musical element, the presentation of the constructed audiovisual product can basically be defined

- a) as “realistic” or “natural(istic)”, when the mediated audio-vision approaches the impression of an equivalent non-mediated event; therefore, the most congruent “realistic” version is the on-stage-performance video including visual elements like the audience, all musicians, technical equipment and so on;
- b) the audiovisual products can be defined as “non-realistic” or “artificial(ized)”, when the visual elements continuously move away from music-related actions to present different types of narrative, situational or simply illustrative additions, which then are more or less unrelated to the underlying musical and textual elements of the song.

Based on the inverse relation of the moving images and the song, in music videos none of the audible musical elements can be produced by any of the visual actions. This leads to the following premises:

- first: in performance-based music videos, usually none of the sound that could possibly be produced by visible actions of playing instruments or singing is actually part of the sound track of the music video;
- second: in rather narrative music videos usually none of the acoustic counterparts of any visible non-musical acts and/or actions – for example a dialogue between two characters – is part of the sound track of the music video.

Despite this material related fact, both these premises can be broken and therefore many recent music videos tend to integrate said variations of additional sounds which then are not part of the original recording of the song but originate from the filmic images.

*Example 1: Sweet Child o' Mine – Guns 'N Roses*

- Watch it here: [https://www.youtube.com/watch?v=JMpK\\_YPpeho](https://www.youtube.com/watch?v=JMpK_YPpeho)

My first example is the rather old music video for *Sweet Child o' Mine* (1988) by Guns N' Roses. It is a typical performance video that shows the band playing the song in a wide rehearsal space changing between colored and black and white images in different angles and shots. In some prior scenes, the musicians and technical staff are shown before the actual performance and because the song is not yet playing, we can hear fragments of voices and the sound of setting up the stage. But most importantly at the end of this prologue we first see a clapperboard and then we see Guns N' Roses guitarist Slash from a bottom-view plugging in his guitar. Simultaneously we hear the feedback of the amp respectively the electric current of both the amp and the guitar followed by the clicking sound of a metronome.

These preceding audiovisual events are not part of the underlying recording as the recording starts with Slash's famous guitar riff. For a while – and maybe only for the initial guitar part – these added technical transmission-related sounds create the impression of a live context regarding time and space as we seem to witness the start of the performance in an audiovisual entity. This short prologue tends to frame the subsequent audiovisual actions as a live documentary of a band performance. However, this cannot really be maintained because on the one hand the preceding visuals also depict other persons than the musicians, on the

other hand there are no more additional music-related sounds. However, the music video for *Sweet Child o' Mine* shows two basic aspects of adding image-based sound elements:

- first, to build an independent audiovisual prologue or epilogue around the pre-existing song;
- second, to add technical sound parts apparently originating from the instruments or the technical equipment, which relate to the production process of the acoustic material.

While the first aspect can be applied to performance videos as well as to rather narrative music videos, the adding of music-related sounds strengthens the impression of a live context by reframing the imitation of a live event as an actual and seemingly causal live event itself. However, the most common type is to include footage of concert audiences and therefore to add audience noises like applauding, screaming, whistling or chanting throughout the music video to pretend that the audiovisual artifact is based on a live performance.

*Example 2: Boom! – System of a Down*

- Watch it here: <https://www.youtube.com/watch?v=bE2r7r7VVic>

A frequent next step is to place the artist's performance in a setting that is not related to a typical location of a musical performance, for example a desert or a crowded public place, which then tends to offer numerous possibilities to add all kinds of sound originating from the actions of the moving images. The music video for *Boom!* (2003) by System of a Down provides a good example as it shows people on the streets protesting against the "War on Iraq" while the song plays in the background. Throughout the music video several protesters can be seen and heard speaking parts of the lyrics into the camera synchronously to the audible lyrics using them as their protest statements.

As a result, on the one hand the musical element of the singing voice is doubled – the most striking aspect is that the singer of the band himself is taking part in the protest and is therefore shown doubling his own voice when he speaks as a protester while he is simultaneously an audible part of the song. On the other hand, the volume of the song is lowered and the added voices are foregrounded, which interferes with the basic materiality of the audio-vision of music videos. Due to these added voices the audible song oscillates between an apparent sound part originating from the moving images and an underlying musical element. Because despite the point of view concept and the changing volume of the song in fact no visible source is provided for the song.

The music video includes all varieties of the combination of the original and the added sound: people speaking during the instrumental part of the song, the foregrounded song without additional sound, the singer doubling his voice part, and the protesters' speaking parts doubling the lyrics.

*Example 3: What Goes Around... Comes Around – Justin Timberlake*

- Watch it here: <https://www.youtube.com/watch?v=TOrnUquxtwA>

Rather narrative music videos usually offer less possibilities of adding music related sounds, however they can include practically every sound that could possibly originate from the actual moving images. There are many narrative and situational music videos that include image-based sounds, like the screeching tires of a car, explosions, gun fire and so on, but not to create a story or a deeper meaning in the audiovisual concept. But then there are examples where image-based sounds are added because the complexity of the aspired audiovisual story cannot be realized solely by the given musical elements and/or the lyrics of the pre-existing song. In these cases, the actual music video can turn into some sort of a movie short with a storyline based on the song's lyrics that is extended by additional filmic sequences with diegetic sounds which are then not related to the song.

The music video to *What Goes Around... Comes Around* (2006) by Justin Timberlake represents a good example for the filmic expansion of the underlying song. The lyrics present aspects of a fatal love story which revolves around the eponymous saying of “what goes around comes around”, however the lyrics are rather generalized to leave room for one's individual interpretation. The visual concept of the music video extends these textual elements by adding several filmic sequences in which the actual song is paused to make room for the added dialogues between the main characters of the visual story.

In the scene, when Justin Timberlake (somehow playing himself), introduces his girlfriend, played by Scarlett Johansson, to his best friend, who is later on having an affair with her, the song is paused at the end of a beat and furthermore at the end of the verse. The inserted dialogue scene provides a broader interpretation to the concept of the audiovisual story, that is, at first based on the lyrics and enhanced by the moving images. Therefore, these added dialogue scenes help to create an audiovisual drama beyond the given lyrics by providing additional textual and visual information about the generalized story, mainly by defining the visual characters and their relationships.

*Example 4: Da Funk – Daft Punk*

- Watch it here: <https://vimeo.com/180840520>

In a further step, the underlying song itself can be transformed into a musical element apparently originating from the moving images, which can result in an equal coexistence of the primary materiality of the song and added image-based sounds. I want to mention just the most prominent type, which is the visual object of the radio as the designated source of the audible song.

The music video for *Da Funk* (1995) by Daft Punk shows a radical way how to implement the adaption of the pre-existing song as a part of the constructed diegetic sound world. The dog-faced man walks the streets of New York with his ghetto blaster constantly playing the song but only in a volume where you can also hear the surrounding street noises and in particular his conversation with a female childhood friend. All sound elements including the song are related to the camera position, so for example, at the end of this scene, when the two characters are shown in a full shot from the opposite side of the street, we can only hear the passing cars but not the song playing on the ghetto blaster.

In *Da Funk* the recording is used as background sound for an unrelated story or a short narrative that is essentially based upon the added urban sounds and in particular on the dialogues. However, in this case the audiovisual concept benefits from the fact, that the song is an instrumental dance track with only little variations of the musical form parts.

Usually in music videos when the pre-recorded song is used in such a diegetic way or when image-based sounds are added to the given musical elements of the song, there is always a mutual process of negotiation between added sounds, musical features and lyrics of the song. Overall the song is still the primary pop cultural product that is presented and advertised with the help of the moving images of the music video, therefore it has to be assured that added image-based sounds do not interfere with the pre-existing inherent context of the song and/or its lyrics.

*Example 5: A Song for the Lovers – Richard Ashcroft*

- Watch it here: <https://www.youtube.com/watch?v=3FVwCm1u8mA>

The music video for *A Song for the Lovers* (2000) by Richard Ashcroft illustrates this exact interference into the materiality of the pre-existing underlying song. The music video depicts Ashcroft sitting around alone in a hotel room listening to his song. The song is presented as a fully imaged-based sound element, because at the beginning of the music video Richard Ashcroft visibly turns on the CD player which then starts to play the song. Basically, in *A Song for the Lovers* the point of view matches with the point of audition leading to a constant change of the volume of the song when the camera follows Ashcroft around the wide and darkened hotel room. This culminates in a short scene when Ashcroft walks into the hallway and away from the source of the music, thus making the song sound distant and dull.

But the most unique audiovisual aspects in this music video are the two scenes when Ashcroft takes the remote control and pauses the music. The camera switches between full shots and medium shots, and the perspective changes from Ashcroft's point of view to a full shot view of Ashcroft from the opposite side of the room. Therefore, the close relation of the point of view camera and Ashcroft's point of audition is cut, and maybe this exact shot creates the impression that somebody else is present. But the overall silence does not really provide a different point of audition.

Especially in contrast to *What Goes Around... Comes Around*, this pausing of the song and therefore the interference in the acoustic materiality of the song does not have the purpose to exclusively make room for image-based sounds that would advance this situational concept. Rather the so to speak "sound of silence" is all there is. At the same time the aspect of this missing of music emphasizes its significance as underlying acoustic materiality for the audiovisual entity.

According to non-mediated forms of audio-vision, the music video for *A Song for the Lovers* illustrates the most extreme way of creating a "natural" or "realistic" audio-vision by turning the song from the inherent and fixed primary artifact of an audiovisual experience (and presentation) into a manipulated and solely musical element of some sort of audio-vision that is not necessarily a music video anymore. In this way, it appears to be only consequent, that the duration of the song in the music video does not match with any other recorded version of the song.

### *Concluding thoughts*

I tried to point out that music videos are characterized by a specific audio-vision that usually does not bring up questions about a possible diegetic status of the music. However, every visible element or object of the moving images holds an acoustic counterpart that could emerge as a parallel overlaying sound in order to create different types of more “natural” or “realistic” audiovisual relations, whether it is to simulate a live context in a performance video or to display the song as an image-based sound element.

Music videos basically serve to advertise the song as an individually and independently existing artifact which can only be purchased as single product without the imagery of the music video. The given music video examples illustrated that the reason for adding image-based sound elements mostly depends on the overall concept of the music video. To sum up some basic aspects:

- in music videos the subsequently added moving images can create different types of “realistic” audiovisual experiences depending on their relation and nearness to the imagination of a non-mediated counterpart; therefore, the impression of a more or less realistic performance is created in relation to the setting, people, technical equipment and so on, or different types of audiovisual narratives are constructed depending on the relationship of moving images, music and lyrics;
- However, the moving images can never achieve the aspired status of being the actual source of any of the sounds or music, which in the end leaves this question to the perception of the audience/ to the eye of the beholder;
- Adding image-based sounds to the already occupied sound track of the music video countervails to its basic material characteristics by interfering in the musical and rhythmical elements and structure of the primary pop cultural artifact;
- but – as I tried to show on the other hand – the audio-vision of music videos also implies its own deconstruction, therefore new forms of audiovisual artifacts can be created, that oscillate between music video and for example short film.