

Music Syncing as Intermedial Translation

Michele Rota

The increasing accessibility of animation tools prompted by the digital age has brought about new categories of audiovisual artifacts requiring appropriate analytical perspectives. On top of actual animation software, there has been a growing use of educational and entertainment software repurposed to create animations, as well as video editing tools to remix existing animated content. These include software for the creation of animated scores (such as Stephen Malinowski's *Music Animation Machine*, or the AI-based platform *Synthesia*), as well as rhythm games (*Guitar Hero*, *Geometry Dash*) and sandbox games (*Minecraft*, *Line rider*). Many music-synced animated videos produced within these digital and online practices feature an existing piece of music as the starting point around which the visuals are shaped in such a way that they do not merely accompany it, but rather transmediate it. This suggests a different way of looking at audiovisual artifacts, one that treats the musical and visual component as source and target text in a practice analog to translation.

Despite being originally defined as a relation between text-based artifacts, translation studies have since developed the notion out of its medium-specificity, and the resulting term *intermedial translation* has proven itself effective in different contexts. My goal with this paper is to further examine the connection between translation studies and intermediality through a rethinking of the notion of intermedial translation. In doing so, I provide an alternative analytical tool with which to examine music visualization, one that is better suited to recognize the deep link that exists between audio and video in terms of both the production and reception of the artifact. I focus on animated videos produced and disseminated online mostly by amateurs through so-called "casual creators," a category developed by Katherine Compton which includes software that provides a

guided, partly automated, creative process.¹ This choice allows for insights into how a wider sample of the population experiences, understands, and reinterprets music through animated visual imagery, as well as how music is remediated within online communities. Aside from this restriction, I purposefully do not define or develop a specific category for the type of artifacts considered as case studies in this research. Depending on which characteristics are taken to be the focus, these audiovisual works can be understood as part of the visual music tradition,² as instances of music remediation, as an expression of fan content (e.g., as Unofficial Music Videos, or U MVs),³ or, as this paper explores, the result of practices of intermedial translation. Rather than defining a category, my goal here is to showcase the potential of intermedial translation as an analytical tool, finding commonalities among several types of audiovisual works with seemingly very different purposes.

Accordingly, this article is structured in two main sections. In the first part I develop the analytical tools: I first briefly contextualize the field of intermediality and go over the history of the term *intermedial translation*, presenting my own understanding of it in comparison with contemporary research. In order to further explore how the concept of translation can be applied to intermediality, and to audiovisual analysis in particular, I then engage with research from both intermedial studies and translation studies: more specifically, I implement Lars Elleström's modality model to discuss the problematic nature of border crossings and contrast it with a functional understanding of translation so as to achieve a satisfactory notion of intermedial translation. Finally, I relate intermedial translation to audiovisual analysis, engaging with literature about music synchronization. In the second part I put this notion to the test, examining a few examples of audio visualization through educational software and video games repurposed as casual creators. In doing so, I aim to showcase intermedial translation as an

1 Katherine Compton, "Casual Creators: Defining a Genre of Autotelic Creativity Support Systems" (PhD diss., University of California Santa Cruz, 2019), 6.

2 Emmanouil Kanellos points out this intersection of visual music, intermedia, and translation: "Visual music can be understood as a sonic composition translated into a visual content, with the elements of the original sonic 'language' being represented visually. This is also known as intermedia." Kanellos "Visual Trends in Contemporary Visual Music Practice," *Body, Space & Technology* 17, vol. 1 (2018): 26.

3 Dana Milstein, "Case Study: Anime Music Videos," in *Music, Sound and Multimedia. From the Live to the Virtual*, ed. Jamie Sexton (Edinburgh: Edinburgh University Press, 2007), 31.

analytical tool capable of making similarities and differences emerge across different categories of audiovisual artifacts. In particular, I focus on the internet phenomenon of creating amateur animated music videos through online applications such as *Geometry Dash* and *Line Rider*. I argue that the role music plays in these practices and the relation established between visuals and soundtrack in the final product differentiate these artifacts from traditional forms of animation. This suggests an alternative interpretation, in which a musical piece acts as source text while the creative process itself is regarded as an act of translation.

Part 1: Developing the Analytical Tools

A Short History of Intermedial Translation

In the context of intermediality—a notion whose semantic field includes many different types of intermedial relations—a broad distinction is between synchronic and diachronic perspectives.⁴ The former, what Siglind Bruhn calls “heteromediality,” has to do with how different aspects of an object affect each other as they are experienced. It can be defined as “the multimodal character of all media and, consequently, the *a priori* mixed character of all conceivable texts.”⁵ In terms of its application to the audiovisual context, the synchronic perspective, or heteromediality, comprises insights on music as accompaniment, such as how the nature of the soundtrack radically influences the audience’s perception of what is being shown on screen. The epitome of this perspective is the notion of “synchresis” developed by Michel Chion, which he describes as “the forging of an immediate and necessary relationship between something one sees and something one hears.”⁶

4 See Lars Elleström, “The Modalities of Media II: An Expanded Model for Understanding Intermedial Relations,” in *Beyond Media Borders, Volume 1: Intermedial Relations Among Multimodal Media*, ed. Lars Elleström (London: Palgrave Macmillan, 2021), 73.

5 Siglind Bruhn, “Penrose, ‘Seeing is Believeing’: Intentionality, Mediation and Comprehension in the Arts,” in *Media Borders, Multimodality and Intermediality*, ed. Lars Elleström (London: Palgrave Macmillan, 2010), 229.

6 Michel Chion, *Audio-Vision: Sound on Screen*, trans. Claudia Gorbman, 2nd ed. (New York: Columbia University Press, 2019), 5 (see also 64–65). See also Diego Garro, “From Sonic Art to Visual Music: Divergences, Convergences, Intersections,” *Organised sound* 17, no. 2 (2012): 103–113; Michael Filimowicz and Jack Stockholm, “Towards a Phenomenology of the Acoustic Image,” *Organized Sound* 15, no. 1 (2010): 5–12; Joanna Bailie, “Film Theory and

The diachronic perspective, on the contrary, is adopted when considering relations of transfer and transformation of meaning from one medium to the other, rather than how different media combine. This is closer to the notion of transmediality, understood as “the general concept that media products and media types can, to some extent, mediate equivalent sensory configurations and represent similar objects.”⁷ Several scholars—such as Andrew Chesterman,⁸ Sandra Naumann,⁹ and Bruhn¹⁰—have focused on specific relations, often between no more than two media. This approach has produced many notions such as musicalization, novelization, ekphrasis, and various types of cross-media adaptation, which have been employed successfully in different fields.

Intermedial translation, the notion that this paper means to adapt and apply to the analysis of audiovisual works, falls within this broad category of diachronic perspectives, being loosely described as a relation between two artifacts belonging to different media types that shares some similarities with text-based translation. The original formulation of the concept is unanimously traced back to Roman Jakobson’s “intersemiotic translation.”¹¹ Starting from the common notion of translation between languages, he developed this term to describe processes where a verbal text is translated into a non-verbal one. The term has been since redefined to encompass translations between any type of media, as the “verbal” status of the origin text seems to be an arbitrary and ultimately impractical limitation. The adjective *intersemiotic*, while it works in the original context of Jakobson’s subdivision between *intralingual*, *interlingual*, and *intersemiotic*, has since been replaced with *intermedial*, to highlight the importance of the materiality and physicality of media within intermedial relations. The resulting

Synchronization,” *CR: The New Centennial Review* 18, no. 2 (2018): 69–74; Kathrin Fahlenbrach, “Aesthetics and Audiovisual Metaphors in Media Perception,” *CLCWeb: Comparative Literature and Culture* 7, no. 4 (2005).

7 Elleström, “The Modalities of Media II,” 74.

8 Andrew Chesterman, “Cross-disciplinary Notes for a Study of Rhythm,” *Adaptation* 12, no. 3 (2019): 271–83.

9 Sandra Naumann, “The Expanded Image: On the Musicalization of the Visual Arts in the Twentieth Century,” in *Audiovisuology, A Reader*, Vol. 2: Essays, ed. Dieter Daniels, Sandra Naumann (Köln: Walther König, 2015), 504–33.

10 Siglind Bruhn, *Musical Ekphrasis: Composers Responding to Poetry and Painting* (Hillsdale: Pendragon Press, 2000).

11 Roman Jakobson, “On Linguistic Aspects of Translation,” in *On Translation*, ed. Reuben Arthur Brower (Cambridge, MA: Harvard University Press, 1959), 232–39.

term “intermedial translation” has thus been employed by scholars such as Vanessa Montesi and Regina Schober.¹²

Montesi’s essay “Translating paintings into dance” attempts to bridge the gap between media studies and translation studies, a goal that my research also shares. It also reads as an excellent guideline on the application of such concepts, and its success itself represents a strong argument toward bringing the two fields together. However, this is achieved for the most part through her case study, while on a theoretical level the term “intermedial translation” is merely understood as a synonym for rather wide categorization tools such as Irina Rajewsky’s “intermedial transposition.”¹³ While she makes use of the analogy to successfully integrate concepts from both intermedial and translation studies, I believe the term’s strong connotations derived from the latter field might prevent it from becoming a catch-all term. Such a role is already covered by more neutral terms such as the already mentioned “intermedial transposition,” or Elleström’s analogous “media transformation.”¹⁴ My position is that “intermedial translation” is not effective as a synonym for this kind of intermedial relations and should rather be given a narrower meaning in relation to the wider class it tries to replace.

Schober’s essay is similarly concerned with application to a case study rather than theoretical analysis, yet it does touch on a few essential points. By describing translation as the “transformation of a pretext of a particular system into another text or system,”¹⁵ it highlights the similarities and even the identity between “intermedial” and “proper” translation—the latter being translation between verbal-based texts. Furthermore, Schober does a good job at stressing the relation between source and target text as a defining aspect of intermedial translation compared to other relations, arguing that “a translation, as an original artwork in itself, is always a modified ‘version’ of the original, containing features quite distinct and divergent from the reference medium.”¹⁶

12 Vanessa Montesi, “Translating Paintings into Dance: Marie Chouinard’s *The Garden of Earthly Delights* and the Challenges Posed to a Verbal-Based Concept of Translation,” *The Journal of Specialised Translation* 35 (2021): 166–85; Regina Schober, “Translating Sounds: Intermedial Exchanges in Amy Lowell’s ‘Stravinsky’s Three Pieces *Grotesques* for String Quartet,’” in Elleström, *Media Borders*, 163–74.

13 Montesi, 169. The reference is to Irina Rajewski, “Border Talks: The Problematic Status of Media Borders in the Current Debate About Intermediality,” in Elleström, *Media Borders*, 51–69.

14 Elleström, “The Modalities of Media II,” 75.

15 Schober, “Translating Sounds,” 165.

16 Schober, 166.

In order to adapt the notion of translation to intermediality, it is however important to also be aware of media specificities and differences between the media to which source and target texts belong. How borders between different media are traced is thus critical to determine the nature of intermedial relations. In the case of intermedial translation applied to audiovisual artifacts, identifying the differences between audio and visual opens up the possibility to clarify how translation actually happens, revealing which elements cause friction to the translation (therefore requiring a degree of transformation) and which are shared across media (making the translation itself possible). In order to do so I will now examine Elleström's "modality model," a theoretical system that categorizes relations between media. This model in particular has been preferred among similar projects for a couple of reasons. It is one of the most recent attempts at organizing intermedial relations at this scale,¹⁷ and it is built upon several previous models, comprising different perspectives within a cohesive system. Furthermore, while it distinguishes media borders clearly, it also leaves space for developing a variety of narrower, sometimes media-specific, relations, which suits the purpose of this article. Finally, a summary of his model will also prove to be a useful starting point to further develop my understanding of intermedial translation by critiquing Elleström's own category of "media translation."

The Modality Model

Any model that focuses too heavily on objective differences, trying to define clear borders between media, can result in too strict a definition, making it difficult to account for the existence of border crossings and relations between media. The modality model proposed by Elleström avoids such essentialist pitfall by realizing that media specificities emerge from the complex interactions of their characteristics, all of which are shared, to a certain degree, across media. As a result, it opens the possibility for intermedial relations while still establishing solid categories of artifacts.

To do so, Elleström arranges the many potential traits of media products by dividing and pinpointing the axes along which media differentiate from one another. By doing this, he provides the building blocks to create catego-

¹⁷ The modality model was first presented organically in Elleström's 2010 article "The Modalities of Media: A Model for Understanding Intermedial Relations" (in Elleström, *Media Borders*, 11–48) and later improved upon in his 2021 article "The Modalities of Media II."

ries with which to circumscribe artifacts that share a similar configuration of characteristics, while still recognizing connections between distinct media. As he explains, “the core of this differentiation consists of setting apart four media modalities that may be helpful for analyzing media products.”¹⁸ These “media modalities” are categories of media traits, defined on the basis of their position on the spectrum from the physicality of the artifact to the interpretive mind that is perceiving it. Specifically, the four modalities that Elleström identifies are: material, spatiotemporal, sensorial, and semiotic. Any media trait falls into one and only one of these four categories and is therefore a “mode” of that specific modality.¹⁹ I quote here a brief summary of the system:

I have argued that there are four media modalities, four types of basic media modes. For something to acquire the function of a media product, it must be *material* in some way, understood as a physical matter or phenomenon. Such a physical existence must be present in space and/or time for it to exist; it needs to have some sort of *spatiotemporal* extension. It must also be perceptible to at least one of our senses, which is to say that a media product has to be *sensorial*. Finally, it must create meaning through signs; it must be *semiotic*. This adds up to the material, spatiotemporal, sensorial and semiotic modalities.²⁰

These four modalities determine what Elleström calls “basic media types,” which are very broad categorizations of artifacts. Each medium shares at least some modality modes with others, but its specific modal configuration identifies it exclusively. When comparing media types, this kind of granular analysis allows for differences as well as shared characteristics to emerge side-by-side.

The basic media types that I analyze in the context of this paper are the two aspects that combine in the audiovisual medium: the soundtrack and the moving images. I do not mean to propose an essentialist outlook that rigidly separates these two fields or to define them as completely independent from each other, but rather to point out overall differences in the way they are perceived. Images and soundtrack affect each other at a deep level, down to how the information is interpreted by the nervous system through

18 Elleström, “The Modalities of Media II,” 8.

19 Elleström, 46.

20 Elleström, 46. Italics in the original.

processes of multisensory integration (“the mutual implication of sound and vision” as demonstrated by the McGurk effect).²¹ Nonetheless, the visual and auditory channels maintain a fairly high degree of separation and allow us to consider each aspect of the audiovisual medium separately. In fact, as exemplified in Chion’s notion of “audiovisual contract,” while synchresis and contaminations between the two are important aspects of perception, “the audiovisual relationship is not natural but a kind of symbolic contract that the audio-viewer enters into, consenting to think of sound and image as forming a single entity.”²²

In order to clarify some peculiarities of intermedial translation as an analytical tool with which to examine musically-synced works, I will now consider the composite medium of audiovisuality through Elleström’s modality model, temporarily treating the two aspects of sound and image as separate media as I discuss each modality.

Two of the four modalities are particularly useful in this context—i.e., the sensorial and spatiotemporal ones. The “sensorial modality” is defined by Elleström as encompassing “the physical and mental acts of perceiving the present interface of the medium through the sense faculties.”²³ Its main modes then mostly correspond to the five basic human senses: seeing, hearing, feeling, tasting, and smelling. In the case of audiovisual works, it is fairly obvious how sound and image are described in relation to this modality: the term *audiovisual* itself separates its two components on the basis of their respective sensorial modes. It is important to note that having two parallel channels of communication affect how easily the spectator is able to receive information from both media simultaneously. While audio and video fundamentally influence each other, the overlap between the two at the level of their sensory perception (e.g., in case of the McGurk effect) is nonetheless limited in nature. This makes it possible for spectators to easily consider both aspects as a unique whole while still being able to recognize their borders fairly accurately. A similar experience has also been studied in relation to another visual medium closely related to music and usually presenting a degree of synchronization: dance. For example, this is what Aasen writes on the subject:

21 See Kevin J. Donnelly, *Occult Aesthetics: Synchronization in Sound Film* (New York: Oxford University Press, 2014), 6 and 25.

22 Chion, *Audio-Vision*, 249n4.

23 Elleström, “The Modalities of Media,” 17.

For the audience, both the object perceived and the sensory modality used to perceive it differ between music and dance. We see the dance, but we hear the music.²⁴

When it comes to the “spatiotemporal modality,” the relation between the two aspects of the audiovisual medium is more complex. The spatiotemporal modality is described through the established model of three-dimensional space (width, height, depth), with time acting as a fourth dimension. Spatiotemporally, sound-based media exist primarily, though not exclusively, in one dimension: time.²⁵ A piece of music, for example, is wholly temporal insofar as it is sequentially interpreted through an act of perception, the temporality of which is tied to the length of the piece itself. If we consider the still images that constitute the frames of an audiovisual artifact, they are characterized by the opposite spatiotemporal modes compared to its soundtrack. While sounds can only manifest through time, images exist in a bidimensional space and are, in themselves, static. Only through cinema they obtain the illusion of movement, and thus a temporal dimension.

There are nonetheless nuances to this schematic opposition. Some otherwise static visual artifacts can in fact present sequential qualities: this is the case for printed texts, which are “conventionally decoded in a fixed sequence, which makes them second-order temporal, so to speak: sequential but not actually temporal, because the physical matter of the media products does not change in time.”²⁶ The same consideration can be extended to images that, while not as rigidly sequential as written texts, adopt visual strategies that draw from different types of notation. For example, in the case of animated scores such as the ones generated to aid the learning or the performance of a musical piece, the animation can be said to simply enhance the sense of movement and direction already implicit in the text contained by each frame. Finally, regardless of the eventual pseudo-tempo-

24 Solveig Aasen “Crossmodal Aesthetics: How Music and Dance can Match,” *The Philosophical Quarterly* 71, no. 2 (2021): 224.

25 It can be, and has been, argued that there is also an inherent spatial element to sound. In the context of music, an interest towards a kind of dramaturgy of space can be traced back to early examples of antiphonal music. The spatial qualities of sound have then been explored more systematically by twentieth-century composers such as Ives, Varèse, Stockhausen and Brant, and still play an essential role in contemporary music production. For a historic analysis of this notion, see Gascia Ouzounian, *Stereophonica: Sound and Space in Science, Technology, and the Arts* (Cambridge, MA: The MIT Press, 2021).

26 Elleström, “The Modalities of Media II,” 49.

ral characteristics of the still images that comprise it, cinema, as its name suggests, is fundamentally temporal. When comparing the soundtrack to its visual counterpart, the shared temporal mode between the two media stands out as the main common element between the two.

When considered together, the sensorial and spatiotemporal modalities help clarify the peculiar nature of audiovisual perception: while they happen simultaneously, thus resulting in a certain degree of overlap, each half can be considered distinct on a theoretical as well as practical level on account of the different sensorial modes that are activated by the video and its soundtrack.

Translation within Intermediality

Modal differences are not only useful in setting up media borders but can also help qualify the types of relations taking place between them. On account of the characteristics showed in the previous section, audio and video can be considered to be fairly dissimilar media types, which poses questions if we are to apply the notion of translation to them. As Elleström makes clear, “given that media types and media borders are of various sorts and have different degrees of stability, it follows that media interrelations are multi-faceted.”²⁷ I find it important to add that this consideration does not outright exclude relations between particular media types: intermedial relations are always possible, what changes is how much the source text needs to be transformed in order for its “meaning”²⁸ to be expressed through a different medium. In other words, artifacts belonging to similar media types can transfer their meaning to each other quite easily, while very dissimilar media require a higher degree of transformation in the process. All of these differences put media and their relations on a spectrum, and therefore do not allow for rigid distinctions to be established.

Despite this, Elleström also proposes a few subcategories of intermedial relations based on modal differences between media. Among these, I find it useful here to discuss his notion of “media translation,” which he contrasts to “media transformation.” In Elleström’s framework, both are understood as forms of transmediality, but the former is intermedial while the latter is

²⁷ Elleström, “The Modalities of Media II,” 71.

²⁸ I am aware that notions such as meaning, content, message etc. are effectively impossible to separate from the artifact’s medium. Here I am using the term in a loose way. In actuality, the translation process consists in recreating a functionally analog artifact, rather than transferring the meaning from one medium to another.

intramedial—²⁹ in other words, *transformation* happens between modally dissimilar media, while *translation* only concerns modally similar media. The core of the inter/intra dichotomy lies in drawing clear distinctions on the basis of modal differences, which are, as I noted previously, on a spectrum. Consequently, I do not believe that these subcategorizations are justified on a theoretical level, and should only be considered if they are found to be useful analytical tools. In the first place, “media transformation” seems to be an appropriate term not only for Elleström’s notion of intermedial transmediality, but transmediality as a whole, on account of the fact that, as he admits, “media differences bring about inevitable transformations.”³⁰ Therefore, to imply that media translation (as understood by Elleström) does not itself entail any form of transformation seems disingenuous.

More concerning is the employment of the term *translation* itself, as Elleström’s notion seems to waste the potential semantic richness inherent in the connection to so called “proper,” that is, verbal-based, translation. Elleström justifies employing the term *translation* out of a desire “to adhere to the common idea that translation involves transfer of cognitive import among similar forms of media, such as translating written verbal language from Chinese to English.”³¹ Such a statement, however, is rooted in a simplistic notion of translation that is now widely considered antiquated. The core bias displayed here has been thoroughly critiqued by scholars such as Maria Tymoczko’s, who links the emphasis on verbal translation to a particularly western-centric understanding of cognition as a whole.³²

Even though Elleström’s “media translation” is clearly an attempt to widen the spectrum of translation to other (similar) media, the effort ends up being misdirected: instead of recognizing the specificities that translation shares with a subset of transmedial processes, the mention of verbal languages is merely exploited as an analogy for media with similar modality modes. If verbal translation is taken as a model without an understanding of what characteristics it actually shares with its intermedial counterpart, the multimodal approach itself is betrayed.

A more promising perspective comes from Klaus Kaindl, who, attempting to integrate multimodality with the notion of translation, disregards aprioristic limitations such as modal similarities and instead concludes:

²⁹ See Elleström, 73–86.

³⁰ Elleström, “The Modalities of Media II,” 75.

³¹ Elleström, 75.

³² Maria Tymoczko, *Enlarging Translation, Empowering Translators* (Manchester: St. Jerome Publishing, 2007).

If we take multimodality seriously, this ultimately means that transfers of texts without language dimension or the concentration on non-language modes of a text are a part of the prototypic field of translation studies.³³

This means, for example, considering the iconic (both visual and audio) elements of a movie as eligible source texts for practices of translation. This kind of approach has not only been called for in theoretical writing, but has also found actual use within translation studies in the practice of audiovisual translation. Here is a brief illustration of its goals and areas of interest:

Audiovisual translation is a blooming subfield of Translation Studies, that includes research areas on practices such as subtitling, dubbing, along with new practices such as fansubbing and crowdsourcing, as well as concerns in accessibility with forms that include subtitling for the deaf and the hard of hearing, and audio description for the blind and the visually impaired.³⁴

Especially in the latter two cases, the translator is moving between media that are far from modally similar.³⁵ Elleström's notion of media translation, then, ends up excluding already established intermedial translation practices such as subtitling for the deaf (transfer from iconic, as well as symbolic, audio elements to visual symbolic elements), and audio description for the blind (transfer from iconic visual to symbolic audio), solely on account of their strong modal differences.

In accord with Kaindl's analysis, I believe intermedial translation's specificity lies in its function and has nothing to do with the dissimilarity of media types or the related degree of transformation that the message goes through in the process of transfer. In fact, this perspective, according to which something is transferred more or less unscathed while other expendable elements are "lost in translation," hides an essentialist view of media products. The message of any text or media product is not produced by a delimited section or aspect of the object but by all of its parts in conjunction; in other words, meaning is influenced by the text as a whole. If this observation is taken to its logical conclusion, it implies that translation does

33 Klaus Kaindl, "Multimodality and Translation," in *The Routledge Handbook of Translation Studies*, ed. Carmen Millán and Francesca Bartrina (London: Routledge, 2013), 266.

34 Vasso Giannakopoulou, "Introduction: Intersemiotic Translation as Adaptation," *Adaptation* 12, no. 3 (2019): 201.

35 This is what Andrew Chesterman calls "media-changing translation;" see Chesterman, "Cross-disciplinary Notes," 274.

not simply transfer the core of the message by losing contextual elements, but—insofar as these elements are indistinguishable from the core ones—it always implies partial transformation of the original message. The reason as to why they are undistinguishable is because what is or isn't essential gets decided by the translator on a case-by-case basis: being first and foremost a mediation practice, translation changes both the superficial as well as the deeper elements of the text that it translates.

This consideration is true for so called “translation proper” (see *supra*), but an intermedial or multimodal perspective makes this the more apparent. In the context of audiovisual works, translation can have very different priorities: in dubbing, the requirement of precise lip sync far outweighs any consideration of literal accuracy, and even in subtitling a lot of thought has to be put into how phrasing affects the timing of a line in relation to what is shown on screen. It is a commonly shared position that this only concerns intermedial instances of translation, while notions of equivalence and reversibility are still championed in authoritative works (such as Eco's)³⁶ as defining characteristics of text-based translation in contrast to its intermedial counterpart. Various researchers have since criticized this position, either indirectly, like Chesterman,³⁷ as well as directly addressing Eco's work, as in the following note by João Queiroz and Pedro Atã:

Eco stresses cases of “adaptations” which are not “translations,” typically because they do not allow an observer to reconstruct the source from the target. We do not consider this criterion (ability to reconstruct the source-sign from the target-sign) to be necessary for considering some communication a translation.³⁸

Conversely, an understanding of translation based on function, and informed by a modal framework such as Elleström's, avoids the pitfalls of the verbal paradigm. Kaindl illustrates this very well:

Such a definition differs from language-centered approaches in two respects.

³⁶ See Umberto Eco, *Experiences in Translation*, trans. Alastair McEwen (Toronto: University of Toronto Press, 2003).

³⁷ For example: “A target text is often said to ‘represent’ (semiotically speaking), or ‘count as,’ its source. But neither ‘representing’ nor ‘counting as’ are necessarily reversible relations.” Chesterman, “Cross-disciplinary Notes,” 280.

³⁸ João Queiroz and Pedro Atã, “Intersemiotic Translation, Cognitive Artefact, and Creativity,” *Adaptation* 12, no. 3 (2019): 311–12n4.

On the one hand, there is no fixed relation to the source text based on similarity, equality, or equivalence and, on the other hand, the role of language in translation does not take centre stage.³⁹

What replaces the source text as the core aspect of this process is “the function of the translation in the target culture, which depends on the actual context of use as well as the expectations and the level of knowledge of the target audience.”⁴⁰ This highlighting of the subjective aspect is crucial, as it fits translation in our contemporary understanding of media in general. Because the meaning of the text itself emerges through a negotiation process,⁴¹ the subjective interpretation of the translator becomes integral to the translation itself: their understanding of the source text dictates the content and form of the target text, as well as its intended function. The relation to the source text is not completely erased, as it persists through the translator’s interpretation of it, but the focus has shifted towards the act of creation of the target text. What is completely cut out of the definition is any reference to language as a hierarchically more significant medium. Kaindl states explicitly that a necessary step to advance media studies as well as translation studies is “to develop translation-relevant analysis instruments for other modes. Linguistic theories as well as image- and music-theoretical approaches are suitable for this purpose.”⁴² This is what I will attempt to do in the following section by linking intermedial translation with the specifically audiovisual practice of music syncing.

Music Syncing

The gist of my proposal regarding music syncing as intermedial translation can be summarized as follows: when the temporal aspect of visuals and soundtrack coincide, in other words when the two are synchronized, this acts as an interpretive bridge that allows for the creation of relations between opposite (or complementary) aspects of the two media within the same artifact. While their opposite sensorial modes entail a great degree of transformation whenever information is transferred from one medium to

39 Kaindl, “A theoretical framework,” 58.

40 Kaindl, 54.

41 On the concept of “negotiation,” see Umberto Eco, *Mouse or Rat? Translation as Negotiation* (London: Phoenix, 2003).

42 Kaindl, “A theoretical framework,” 65.

the other, their temporality forms the common ground that makes establishing this transfer possible in the first place.⁴³

Any audiovisual product, insofar as it displays both audio and visuals, is, at least in theory, eligible to be analyzed through the lens of intermedial translation; however, instances of actual correlation between the action on screen and the musical events determine whether the potentiality of this specific interpretation actually takes place. As Donnelly explains in his book *Occult Aesthetics*,

as time-based art forms, film and music have many fundamental attributes in common, despite clear differences. Isomorphic structures on large and small scales, and patterns of build-up, tension, and release are most evident in both... Film, both in terms of its conceptualization and practical production, embraces sound and image analogues (in terms of rhythm, both literally and metaphorically), “Mickeymousing” in music, or gesture and camera movement being matched by or to the “sweep” of music, to name but a few instances.⁴⁴

These correlations between sound and image can be analyzed through the lens of a synchronic perspective, showing how the two aspects affect each other as they are perceived. However, in accord to Elleström’s position that “no media products exist that cannot be treated in terms of diachronicity without some profit,”⁴⁵ I argue that, whenever an audiovisual object is experienced, it always has the potential of being read in the diachronic perspective as well, as an instance of intermedial translation. In other words, the visuals can be interpreted as translating the soundtrack, or vice versa. Whether the visuals or the soundtrack are considered to be source text varies on a case-by-case basis and is dependent on both the specificities of the artifact as well as the perceiver’s context and their inferences regarding the audiovisual object itself. In the examples that I will examine in the following pages, the direction usually goes from music to video, as intermedial translation often overlap with instances of music remediation, where an existing piece of music is incorporated into an audiovisual work. This

⁴³ Discussing audiovisual synchronization, Kevin Donnelly sums up the complex nature of the medium as follows: “One of its open secrets is the separation of sound and image and their union through mechanical (and increasingly digital) synchronization.” Donnelly, *Occult Aesthetics*, 3.

⁴⁴ Donnelly, 5.

⁴⁵ Elleström, “The Modalities of Media II,” 73.

does not imply the primacy of one medium over the other,⁴⁶ but merely that it is possible to interpret an audiovisual work as a translation between its soundtrack and its visuals. According to this perspective, translation isn't just dependent on its perceived function, but is itself a potential function of the target text. Reading a media product as a translation is a strategy that the perceiver can employ to better understand the target text on account of its relation to a (perceived) pre-existent source text, or rather to understand both target and source texts on account of what their relation helps to highlight. This reading is indebted to Elleström's own understanding of intermediality as a whole.⁴⁷

The perception of specific traits in the target text and their similarity to aspects of the source text is entirely dependent on the interpreter's subjectivity. Consequently, insofar as translation is not only a "text-processing activity" performed by the translator, but also and more importantly an interpretive practice employed by the perceiver, what translation ultimately means ends up being decided by the audience, based on their context. Similarly, Cook reminds us that "inter-media relationships are not static but may change from moment to moment, and that they are not simply intrinsic to 'the IMM itself' [instance of multimedia], so to speak, but may depend also upon the orientation of the recipient,"⁴⁸ that is to say, the subjective interpretation of the perceiver.

Similarly to what happens with interlingual translation, the same audiovisual artifact, presenting the source text as well as its translation, will be interpreted differently based on the reader's level of proficiency with and ease of access to the source and target language, ranging from virtually disregarding one of the two texts, to a bilingual level, in which case the translation is seen as actualization of one particular interpretive perspective rather than a tool meant to make the source text accessible. Naturally, intermedial translation always has to compete with different potential interpretations, such as regarding the soundtrack as mere accompaniment for the visuals rather than a media product in its own right. Axel Englund

46 If this were to be the case, it would fall among the "hegemonic models in multimedia theory—the idea that one medium must be primary and others subordinate," a notion that Nicholas Cook extensively criticizes in his research. Cook, *Analyzing Musical Multimedia* (Oxford: Clarendon Press, 1998), 117.

47 I refer here to statements such as: "Intermediality is an analytical angle that can be used successfully for unravelling some of the complexities of all kinds of communication." Elleström, "The Modalities of Media II," 4.

48 Cook, *Analyzing Musical Multimedia*, 113.

sums up effectively this understanding of intermedial translation as an analytical perspective:

The emphasis on the interpretative activity is important: what we are dealing with is less an inherent quality exhibited by intermedial works than a fertile strategy for approaching these works.⁴⁹

One approach doesn't exclude the opposite one, and as long as it is understood as a perspective rather than a categorization tool, the notion of intermedial translation developed so far can be applied to any class of media. In the next section of the article I will test it on a range of contemporary practices of music remediation that have developed mostly online through the past decades. However, before doing so, it is useful to very briefly look at different types of music synchronization in order to establish a few guidelines. A fairly direct approach to this is to examine the degree of synchronization that takes place in an audiovisual artifact. This is how Donnelly proceeds:

Broadly, there are three types of sound and image relation. These run on a spectrum from tight synchronization at one pole and total asynchrony at the other. In between, there is the possibility of plesiochrony, where sound and image are vaguely fitting in synchrony, but lack precise reference points.⁵⁰

While the distinction between tight and loose synchronization is an important and useful way to differentiate audiovisual works, this approach alone is inadequate to specify the quality of synchronization, meaning in what way images and sounds actually correlate to each other. Without introducing a multiplicity of categories, it is possible to at least recognize two different paradigms, which respond to different functions. Media scholar James Tobias qualifies synchronization by identifying two antithetical model: "phrasal" and "rhythmic."⁵¹ These two styles of synchronization can be regarded as the two opposite paradigms at the core of visual-to-music translation. When interpreted this way, any actual work falls somewhere in between on a spectrum defined by these models.

49 Axel Englund, "Intermedial Topography and Metaphorical Interaction," in Elleström, *Media Borders*, 70.

50 Donnelly, *Occult Aesthetics*, 31.

51 James Tobias, *Sync: Stylistics of Hieroglyphic Time* (Philadelphia: Temple University Press, 2010), 99.

On one side, an extremely schematic frame which transposes music on the basis of strict parameters: this is quite similar to, and partially derived from, the system developed in the western tradition to record music in written form. According to this paradigm, music is conceived spatially in terms of *higher* and *lower* pitches, which can be reproduced as easily on film as they are on a pentagram. I have chosen to refer to this first kind as *notational* rather than “phrasal,” as its main characteristic seems to be the almost isomorphic mapping of musical information. On the opposite side, a looser but potentially more expressive interpretation, which has its roots in dance rather than written notation. In place of a heavily codified transposition of musical elements into visual ones, dance subjectively illustrates music’s emotional content through abstract movements and pantomime.⁵² Because of this similarity and the centrality of the dance model in so many descriptions of audiovisual works that present these characteristics, I refer to this type of synchronization as the *choreographic paradigm*, since “rhythmic” is too vague a descriptor. On a theoretical level, distinguishing the two paradigms proves useful, as it provides a tool for the analysis of audiovisual works. However, in practice most animations are informed by both models and showcase a mix of characteristics.

In the second half of this paper I am going to apply the analytical tool of intermedial translation as I have developed it in the previous sections. To recap briefly: I have explained the use of the term *intermedial* as it is a relation between artifacts belonging to two different media, and the word *translation* to highlight the analogy with characteristics developed by translation studies, in particular as they pertain to the notion of translation as function. Whenever an artifact is examined as an instance of intermedial translation, the process that led to its creation is likened to the translator’s, who necessarily adds a subjective interpretive layer to the transposition from the source to the target artifact.

This analytical tool can be applied to the analysis of audiovisual artifacts to examine the relation between their audio and video components. According to this perspective, the synchronization of soundtrack and images is understood in terms of one aspect translating the other. This can be es-

52 For literature about the congruence of music and dance, see Carol L. Krumhansl and Diana Lynn Schenck, “Can Dance Reflect the Structural and Expressive Qualities of Music? A Perceptual Experiment on Balanchine’s Choreography of Mozart’s Divertimento N. 15,” *Musicae Scientiae* 1, no. 1 (1997): 63–85; see also Robert W. Mitchell and Matthew C. Gallaher, “Embodying Music: Matching Music and Dance in Memory,” *Music Perception* 19, no. 1 (2001): 65–85.

pecially fruitful in instances of music remediation through audiovisual artifacts, which are therefore understood as analogue to a book with original and translated text put side by side. The analogy with translation also helps to establish a spectrum of possibilities in terms of how the soundtrack is translated into a visual interpretation. We can think of the two paradigms suggested by Tobias as different degrees of literality: the notational paradigm representing the more “literal” and somewhat objective translation, and the choreographic paradigm the more subjective interpretation.

Part 2: Animated Scores

The digital age, especially with the wide adoption of digital technologies in the 1990s and early 2000s, saw a flourishing of music-to-visual translation software and practices. These include innovative ideas as well as a significant re-mediation of old media by the new digital tools. Furthermore, the growing accessibility of digital tools greatly accelerated the already ongoing process of blurring the lines between so called high art and low art. Accordingly, in this second half I intend to explore practices at the intersection of art, entertainment, and education. An exhaustive overview is of course outside the scope of this article: my goal here is rather to touch upon a variety of software to demonstrate the width of practices that can be productively grouped together and compared by applying an intermedial translation perspective.

Music Animation Machine

The first example I cover is Stephen Malinowski’s *Music Animation Machine*, a pioneering visualization software. It was first developed in the mid-1980s, but progressively modified and perfected through the past three decades. Matt Woolman’s book *Sonic Graphics/Seeing Sound* offers a good description of it:

This piece of software by Stephen Malinowski makes animated graphical scores of musical performances. The scores are specifically designed to be viewed on videotape or computer screen while listening to the music. Stephen Malinowski’s inspiration comes from the visual theories of Paul Klee, and a desire to extract conventional musical scores from their static page-frames. They contain much of the information found in a conventional score, but

display it in a way that can be understood intuitively by anyone, including children.

Each note is represented by a colored bar and different colors denote different instruments or voices, thematic material or tonality. ... The bars scroll across the screen from right to left as the piece plays, and each bar brightens as its note sounds to provide a visual marker for the viewer. The vertical position of the bar corresponds to the pitch: higher notes are higher on the screen, lower notes are lower. The horizontal position indicates the note's timing in relation to the other notes.⁵³

The denomination “animated graphical score” already characterizes this music-to-video translation as pertaining to the strict notational paradigm. In fact, when documenting the various stages of his process, Malinowski notes that his initial goal was to update musical notation, in order to make it more accessible to listeners regardless of their level of musical competence. Here is what he writes:

I made my first graphical score because I wanted to eliminate some of the things that made it hard to follow a conventional, symbolic score in real-time: a complicated mapping of pitch to position (multiple staves, multiple clefs, and sometimes multiple transpositions), and the need to remember which staff corresponds to which instrument (these are labeled at the first page of a movement, but usually not on subsequent pages, and only within a page when there's a change of instrument on a staff), and a symbolic system for showing timing.⁵⁴

Timing is a particularly important feature, as it arguably represents the least intuitive aspect of traditional notation. The choice to visualize it as movement along the horizontal axis is the logical adaptation of the left-to-right flow that is already embodied in the pentagram as well as verbal written text. The resulting score is a mostly mathematical representation, in the form of a diagram. At its core, it converts the basic pitch and rhythm by placing nodes on a graph (the vertical axis recording pitch, the horizontal one time), in such a way as to replicate traditional musical notation

53 Matt Woolman, *Sonic Graphics: Seeing Sound* (London: Thames & Hudson, 2000), 10.

54 Stephen Malinowski, “What am I doing?” on his personal website *musanim.com*, accessed October 17, 2023, <http://www.musanim.com/WhatAmIDoing/>.

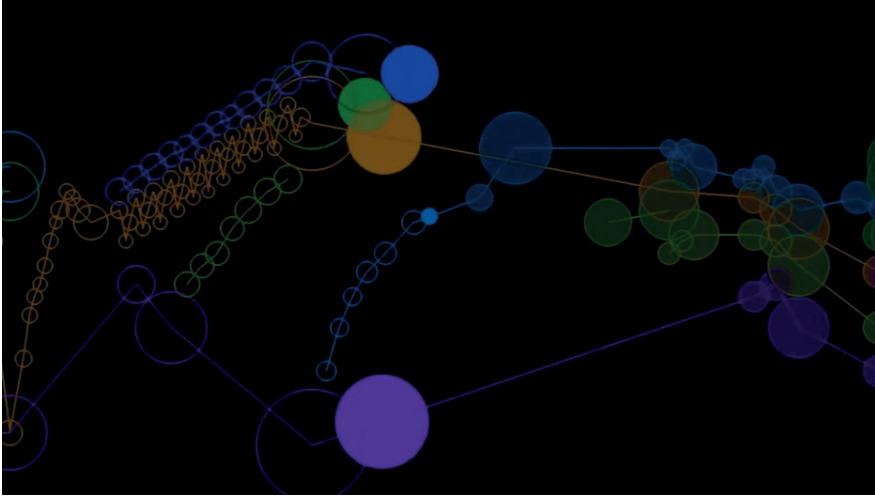


Fig. 1 – Still frame from Malinowski's first animated score, visualizing Debussy's *First Arabesque* (see note 55)

in a more simplified form which mostly disregard key, bar lines, etc. Other information such as volume or which instrument is playing is signaled through the color and shape of the nodes.

As a result of this very rigid structure, most of the process can be automatized (the software takes information from a MIDI file as basis for its visualization). There are however many musical elements which are not unambiguously transposed into visuals: tone, phrasing, and even which layer should a note occupy based on its perceived role as a melodic, harmonic, or rhythmic element. All of these require a translator's interpretive process, and result in an artifact that is not merely an automatized transposition of audio information into a visual medium. Recounting his first actual visualization,⁵⁵ Malinowski remarks: "Though I didn't realize it at the time, what I'd begun doing was *embodying aspects of my perception of the music in the animation*."⁵⁶

On this note, his own considerations on the *Music Animation Machine* in relation to other media are illuminating. The comparison to dance is particularly interesting. On one hand he states that, strictly speaking, his

55 musanim, "Debussy, First Arabesque," YouTube, animated score, uploaded on April 8, 2010, <https://www.youtube.com/watch?v=fELE-LFep9w>.

56 Malinowski, "What am I doing?"

animations have little to do with forms like ballet. This should not be surprising, given that it represents the opposite paradigm to the notational one at the core of Malinowski's entire project. On the other hand, however, he remarks that there are in fact similarities with the process of creating a choreography, even if only metaphorical ones:

I was, in a way, "choreographing"—not choreographing the animation, but choreographing the audience's attention—ideally, helping them attend to the things they'd attend to if the[y] "knew what to look for."⁵⁷

It is remarkable that Malinowski manages to look at the more subjective dance paradigm through the lens of aiding the spectator's understanding of the music, somehow blending the two aspects together in his animated scores.

Music Visualization and Rhythm Games

While technically a music visualization software, I argued that Malinowski's *Music Animation Machine* constitutes a special case by virtue of the role his own subjective interpretation plays in the production process. In contrast with this balance of rigid framework and personal sensibility, I want to briefly touch upon entirely or mostly automated audio visualizers.

Generally speaking, the high number of software that allow for music visualization fall somewhere between entertainment and educational purposes. On one side, the default media player installed on most PCs since the early 1990s, *Windows Media Player* (1991), provides a basic counterpart to the music by transposing it in real time into abstract visuals. This was initially done for the purpose of showcasing the technical capabilities of the operating system but has since become an almost archetypical example of a distinct digital aesthetic. Despite the absolute degree of automation, some visualizations of *Windows Media Player* (and similar software) process the audio input algorithms in such a way as to create a sort of choreography. As touched upon earlier in this paper, the crucial element to intermedial translation isn't the production process itself but how the resulting artifact is interpreted by the viewer. As a consequence of all the visual noise, these audio representations don't provide the viewer with an easily accessible notation, and rather tend towards the choreographic paradigm.

⁵⁷ Malinowski, "What am I doing?"



Fig. 2 – Still frame from a video of Debussy's *First Arabesque* with *Synthesia* visualizations. Rousseau, "Debussy - Arabesque No. 1," YouTube, uploaded on May 28, 2018, <https://www.youtube.com/watch?v=cVYH-7QGE-A>.

On the opposite end, *Synthesia* (2006) is one of many software developed specifically for the purpose of teaching music by offering alternative visual notation. Similarly to Malinowski's *Music Animation Machine*, it draws information directly from a MIDI file, but its visualization is severely limited. It is conceived as a piano keyboard trainer, which is why the information displayed is narrowed down to the basic elements of pitch and timing, and the visualization itself is tied to a virtual keyboard rather than an expressive presentation of the music. Consistent with the software's main function, the resulting artifacts can be clearly described as embodying the notational paradigm. Despite the limited aesthetic tools, performances of piano pieces accompanied by the *Synthesia* visualization of the corresponding MIDI files represents a fairly popular genre of YouTube videos which attract an audience that enjoys the visuals beyond their educational purpose.

A third example is offered by rhythm games, a subgenre of action video games, usually music-themed or dance-themed, in which the player is required to perform a long series of rhythmic inputs with relatively precise timing. The peculiar nature of these software provides interesting insights into one way to visualize music effectively. The soundtrack accompanying each level is an indispensable frame of reference that makes the execution and memorization of tasks significantly easier and more enjoyable. Because



Figure 3 – “Note Highway” from *Guitar Hero* (left) and its 3D adaptation in *Beat Saber* (right).

the goal of rhythm games is to guide the player into performing specific actions, their music visualization must be not only extremely precise but also immediately readable by any type of audience. Each significant musical element that is supposed to coincide with a user input has to also be accompanied by a clearly defined visual element. Similarly to a conductor’s gesture, these elements can be said to interpret the music, while at the same time they act as cues for the player. This synthesis of different roles played by the soundtrack in relation to the gameplay is fairly common in video games.⁵⁸

The way the visualization is displayed in games such as the *Guitar Hero* series shares many similarities with Malinowski’s as well as other alternative forms of musical notation.⁵⁹ In the case of rhythm games, differences in pitch are sometimes disregarded, or rather simplified (for example, by grouping all possible notes in four or five different inputs), while the rhythmic aspect is standardized and made easier to read. The visualization consists of the so called “note highway”:⁶⁰ a vertical scrolling bar on which colored circles flow downward until they intersect a horizontal line at the

⁵⁸ “As the music and gameplay come together in games, these simultaneous synchronous domains of play converge to create a complex nexus.” Tim Summers, *Understanding Video Game Music* (Cambridge: Cambridge University Press, 2016), 6. See also Kiri Miller, *Playing Along: Digital Games, YouTube, and Virtual Performance* (New York: Oxford University Press, 2012); Roger Moseley, *Keys to Play: Music as a Ludic Medium from Apollo to Nintendo* (Oakland: University of California Press, 2016).

⁵⁹ James Tobias also points out similarities between rhythm games, audio visualizers, and the wider category of visual music: “The popular *Guitar Hero* series and the automatic frequency and rhythm analyzer, which come standard with digital music players, are not the end of the line; indeed, they are more simply statements of the meanings of ‘visual music’ as a problem of everyday hermeneutics for or stylization of uses of time as material.” Tobias, *Sync*, 83.

⁶⁰ Originally developed by game developer Konami for *GuitarFreaks* (1998).

bottom, which signals the player the exact timing for the corresponding input. This visualization proves to be effective and has been replicated in a variety of musical games and minigames. This same visual solution also informs educational software such as the already mentioned *Synthesia*.

In all these cases the reading as translation is made possible by the extremely tight synchronization, which in turn results from the automatic nature of the visualization process. Despite the lack of, or the limited nature of, subjective inputs in the translation process, I have argued that these animated videos can nonetheless be read as instances of intermedial translation by the audience, as long as the visuals are perceived to interpret the soundtrack or vice versa.

Casual Creators

Having presented a few examples of mostly automated music visualization, the next step is to look at instances where a partly automated process is employed by individual users to create their own intermedial translation of pre-existing musical pieces. The mention of rhythm games in particular can serve as an ideal starting point to introduce the notion of *casual creators*. This term refers to a class of software which Katherine Compton defines as follows:

Casual creators are interactive systems that encourage and privilege fast, confident, and pleasurable exploration of a rich possibility space; leading the user to experience feelings of creativity, both in a sense of ownership of what they make, and of their participation in an environment of creativity.⁶¹

The core element is the promotion of a creative process founded on the playful exploration of an interesting artifact rather than creation out of nothing. The process itself depends on meaningful choices made by the user but is also partially assisted by the casual creator, so that it “produces artifacts within a limited-yet-meaningful domain space, enabling automation and support, both passive (encoded into the domain model and system constraints) and active (responding to user actions).”⁶² In doing so, the user is provided with a framework that makes the creative process less demand-

61 Compton, “Casual Creators,” 6.

62 Compton, 6.

ing and therefore generally more enjoyable compared to the workflow enabled by professional artistic tools.

In the present section I am interested in exploring the intersection of casual creators and visual music. In particular, I wish to examine what happens when the process of creating artifacts through a casual creator is intrinsically linked to a musical piece, chosen at the beginning of the creative process itself as accompaniment for the final product. Because a casual creator is usually embodied by a virtual or metaphorical environment which is subsequently explored by the user, I suggest that, whenever music is involved, the exploratory process of interacting with this environment also becomes an exploration of the musical artifact as well.

Whenever they are equipped with an in-game level editor, rhythm games can be repurposed as casual creators, as they typically present a degree of automatization in their visualization of music even when they allow users to create their own charts. While technically outside of the category of rhythm games, many of the same characteristics are also shared by genres such as auto-run platformers (games where the player character keeps moving forwards). As a general rule, in fact, any time the gameplay requires the player to execute actions with a timing that is not tied to random elements—which means it can be designed to be rhythmical—the soundtrack can be employed as a way to elevate the experience, becoming integral to the player's interaction with the game. This is what Donnelly defines as “music-led asynchrony,” one of four types of music synchronization in video games as he classifies them in his research.⁶³

For instance, in *Geometry Dash* (classified as a musical platforming game) music syncing is the norm in both main levels as well as user-generated ones. The game launched in 2013 as a mobile-only game. In each level, the player controls the vertical movement of the character as it moves automatically at constant speed from left to right. Once again, the notational paradigm plays an important role. What differentiates *Geometry Dash* from other auto-run and rhythm games with in-game editors mostly boils down to the surprisingly wide variety of aesthetic choices available to the user, which results in visually unique experiences for players and spectators. Over the years, the addition of more and more editing tools, as well

63 “With games that are timebound, music can often appear to set the time for gameplay.” Kevin J. Donnelly, “The triple lock of synchronization,” in *The Cambridge Companion to Video Game Music*, ed. Melanie Fritsch and Tim Summers (Cambridge: Cambridge University Press, 2021), 104.

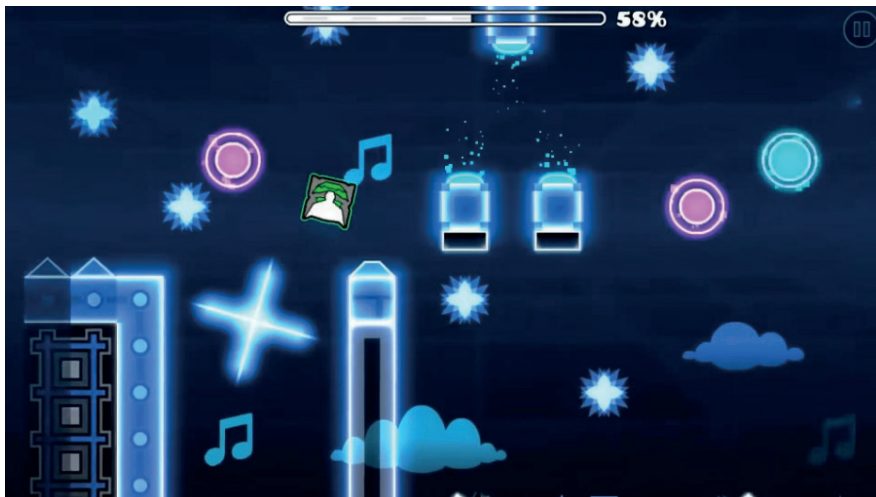


Fig. 4 – Music visualization in *Geometry Dash*. Still frame.

as the ability to customize backgrounds, empowered *Geometry Dash* level creators to complement gameplay with their personal interpretation of the musical pieces they employed as soundtrack.

When designing levels, the choice of which and how many musical elements to sync with a player's action and the leniency in terms of the timing window for each input all entail gameplay decision alongside the process of visualizing music. I argue that these constraints act as limitations as well as guidelines: in other words, they reduce the possibility space, which is a typical aspect of casual creators. This in turn supports the user's creativity by offering a framework of relatively clear directions.⁶⁴ Furthermore, syncing each visual element is not as straightforward in *Geometry Dash* as it is in traditional rhythm games, because of the nature of the game itself: each obstacle does not correspond to a single input, which results in less tight synchronization from the player perspective.⁶⁵ However, because most levels

64 Because of its tight relation to gameplay, music synchronization in rhythm games can be said to display a heightened version of what Donnelly describes about film music: "Points of synchronization constitute something of a repose, a default position of normality that furnishes moments of comfort in a potentially threatening environment that can be overwrought with sound and image stimuli." Donnelly, *Occult Aesthetics*, 73.

65 It is important to note that synchronization, however important in rhythm games, is only one aspect of the video game player's relation with game music. "Constructed by the gamers' listening activity, the sonic experience of a game is conditioned by an individu-

are also recorded in video form as they are completed, the constant speed of both backgrounds and foreground objects in combination with the music creates audiovisual artifacts that can easily be interpreted as instances of intermedial translation by the audience. Through this process of recording gameplay and sharing it online, games like *Geometry Dash* become alternative tools for the creation of abstract animated music videos.⁶⁶

These artifacts can be categorized as a particular type of animation, usually referred to as “machinima.” Johnson and Petit, who co-wrote one of the few books dedicated to this discipline, define it as “the process of capturing and constructing images within virtual environments to tell a story through iconic representation in various forms and genres.”⁶⁷ While technically machinima refers to movies made with any kind of real-time computer graphics engine, the term is usually restricted to animation created within video games world. In doing so, machinima artists sidestep the intended purpose of games and turn them into tools for cinematic production. While not mentioning virtual filmmaking directly, Compton also comments on this same process through the lens of casual creators:

A few systems were not intended as casual creators (e.g. traditional games or professional creativity tools), but were picked up by users who played counter to the expectations of the system, hacked in new features, and adapted these systems into being casual creators.⁶⁸

A demonstration of the wide variety of games that can be used to achieve loosely abstract non-narrative animation is provided by content creator Mark Robbins, who runs the YouTube channel DoodleChaos. His catalog of videos comprises mostly music synced animation made through vari-

al’s acquired knowledge (from a historical, intermedial, and videoludic perspective) as well as his or her perception, and involvement with the audiovisual components of the games.” Guillaume Roux-Girard, “Sound and Videoludic Experience,” in *The Oxford Handbook of Interactive Audio*, ed. Karen Collins, Bill Kapralos, Holly Tessler (Oxford: Oxford University Press, 2014), 143.

⁶⁶ Music then plays a role in every step of the process, as a guideline in level creation, as a gameplay mechanic for the player, and as a soundtrack for the viewer. On the latter aspect, see also Sandra K. Marshall and Annabel J. Cohen, “Effects of Musical Soundtracks on Attitudes Toward Animated Geometric Figures,” *Music Perception* 6, no. 1 (1988): 95–112.

⁶⁷ Phylis Johnson and Donald Petit, *Machinima: The Art and Practice of Virtual Filmmaking* (Jefferson: McFarland & Company, 2012), 4.

⁶⁸ Compton, “Casual Creators,” 7.

ous games, including *MineCraft*,⁶⁹ *Planet Coaster*,⁷⁰ *TrackMania*,⁷¹ as well as musical performances recreated in physics simulators.⁷² All these different audiovisual works fit the category of machinima animations, each software acting as a casual creator in which the synchronization aspect constitutes a framework for the creative process. While Robbins's videos mostly aim to be entertaining rather than serious artistic statements, they nonetheless demonstrate the author's subjective interpretation of the music which accompanies them. In doing so, they also constitute instances of music remediation, communicating to an audience a unique perspective on already established musical artifacts. Because of the DIY character of these practices, the audiovisual works produced represent an invaluable source of information on how a large number (in comparison to the relatively narrow category of professional artists) of individuals express their own subjective experience of music through visual means.

Line Rider

Having looked at visualization software with different degrees of human involvement, as well as traditional games used as cinematic tools, it is now possible to examine a case that sits somewhere in between the two categories.

Line Rider was originally popularized as a game, published on flash games websites, despite lacking many of the core features that would define it as such (e.g., a clear goal, a score system, fail conditions and such). On the other hand, it has been *de facto* repurposed as a tool for the creation of animated music videos, even if it offers limited possibilities compared to actual graphic design and animation software. It is described as a "toy" by its author, highlighting the entertainment aspect. Technically it falls into the category of sandbox games, which greatly overlaps with casual creators. Compton recognizes it as a casual creator and describes it as follows:

69 DoodleChaos [Mark Robbins], "I Synchronized my Minecraft World to Music," YouTube video, uploaded on September 8, 2021, <https://www.youtube.com/watch?v=Y88ezQ8b-pfE>.

70 DoodleChaos, "Music Synchronized Roller Coaster (Front Seat POV)," YouTube video, uploaded on January 24, 2020, <https://www.youtube.com/watch?v=ElrDJO6XemA>.

71 DoodleChaos, "Ride of the Valkyries – TrackMania," YouTube video, uploaded on January 11, 2019, <https://www.youtube.com/watch?v=v6fxsvQxV8g>.

72 DoodleChaos, "Visualizing Piano with Bouncing Balls (Chopin – Nocturne op.9 No.2)," YouTube video, uploaded on May 27, 2020, <https://www.youtube.com/watch?v=tWwTIooB-HKI>.

Line Rider, a 2006 online Flash game by then-student Boštjan Čadež, is a casual creator where the user draws lines, which are then used as a sledding hill by a tiny physics simulated person on a sled. Users began creating intricate and wildly creative levels where the sledder's fall would synchronize with music, staging user-led contests and collecting and archiving levels.⁷³

In the years after its release, *Line Rider* became quite popular by browser games standards. Alongside the many casual players, it also spawned a smaller committed community of creators who mastered the limited tools of the software turning it into what could arguably be considered a distinct artistic medium.⁷⁴ The first decade mostly saw *Line Rider* trackmakers learning to manipulate the physics engine, exploiting glitches, or drawing painstakingly intricate backgrounds for the rider to traverse, with the main goal to display their virtuosity in the context of a fairly competitive environment rather than expressing artistic intent.⁷⁵

A big shift occurred around the mid-2010s, when the viral success of a few *Line Rider* videos accelerated the already growing tendency towards music synchronization within the community. These groundbreaking tracks are fairly different from each other: “Ragdoll,”⁷⁶ a short comedic driven music synced track created in 2016 by David Lu (aka Conundrumer) was one the first *Line Rider* tracks to employ relatively tight synchronization. The next year saw “This Will Destroy You,”⁷⁷ by Bevibel Harvey (aka Rabid Squirrel), a monumental hour-long track named after the self-titled album of a post-rock band, which demonstrated the medium's capacity for a wide range of emotional and artistic expression through the use of simple visual analogues that subjectively interpreted the soundtrack.⁷⁸ Finally, the same year,

73 Compton, “Casual Creators,” 250–51.

74 This claim relies on what Elleström calls the “contextual qualifying aspect,” which “involves forming media types on the grounds of historically and geographically determined practices, discourses and conventions. We tend to think about a media type as a cluster of media products that one begins to use in a certain way, or gain certain qualities, at a certain time and in a certain cultural and social context.” Elleström, “The Modalities of Media II,” 60.

75 See *Line Rider* Review, “A History of *Line Rider* [Lecture],” Youtube video, uploaded on Apr 22, 2019, https://www.youtube.com/watch?v=n_kwIPDWSO8.

76 Conundrumer [David Lu], “*Line Rider* – Ragdoll,” YouTube video, uploaded on February 14, 2016, <https://www.youtube.com/watch?v=LMgnZYpbM-M>.

77 Rabid Squirrel/Bevibel Harvey, “This Will Destroy You: A *Line Rider* Feature Film,” YouTube video, uploaded on July 23, 2017, <https://www.youtube.com/watch?v=qasxqKScOfY>.

78 While the specific tools are different, the basic process and underlying perceptual rules are similar to what has already been studied in relation to audiovisuality as a whole. For

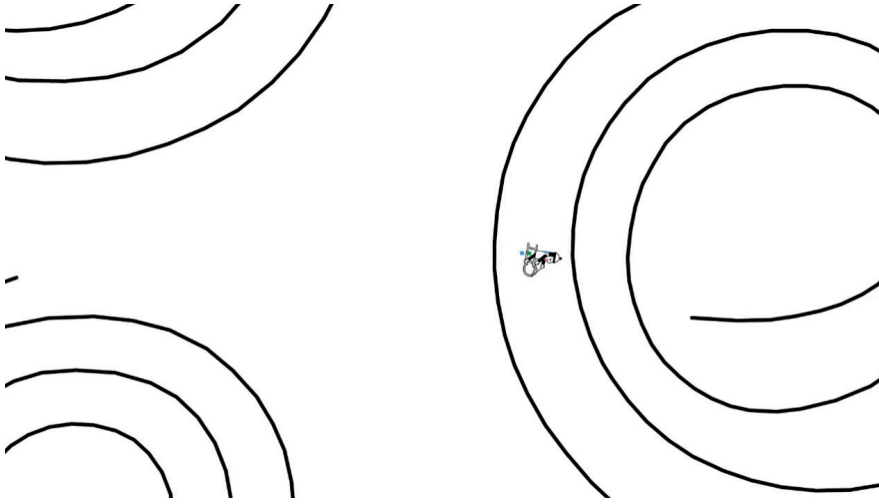


Fig. 5 – Still frame from Beviel Harvey's *This Will Destroy You* (see note 77).

the already mentioned YouTube creator DoodleChaos published his first track, “Mountain King,”⁷⁹ a light-hearted visualization of Grieg’s famous piece which quickly became the most viewed *Line Rider* video of all time.

While the previous technical virtuosity-driven tradition had produced works that were appraised based on parameters virtually inaccessible to anyone outside of the small *Line Rider* community itself, these innovative pieces focused on entertainment and artistic qualities which could be appreciated by a wider audience, building their visuals on relatively simple yet expressive graphic elements on top of tight music synchronization. In the following years, more and more trackmakers began following this new direction. From that point onward, music became *de facto* intrinsic to the *Line Rider* creative process. As I have argued earlier, in such cases the interaction with the casual creators also becomes a way to interpret the musical piece that guides the process. According to Compton:

example: “When sound and image on-screen are synchronized, we associate the two, perceiving them as a single entity, or at least as directly and intimately related. Once this takes place, other attributes of sound become attributed to the visual object too.” Donnelly, *Occult Aesthetics*, 24–25.

⁷⁹ DoodleChaos, “Line Rider – Mountain King,” YouTube video, uploaded on October 7, 2017, <https://www.youtube.com/watch?v=RIz3klPET3o>.

The experience of a casual creator is in using the tool as a probe to explore the possibility space, and understanding one's movement through that space by engaging with the artifact itself, not just an abstraction in the user's mind.⁸⁰

The two-dimensional infinite white canvas that constitutes *Line Rider*'s virtual environment therefore represents the visual counterpart that allows for a unique way of interpreting musical artifacts. In other words, whenever it is used as a tool to express music visually, the exploration of the *Line Rider* engine also becomes an exploration, through its visual translation, of the music piece itself. This notion is at the core of my understanding of practices of intermedial translation within casual creators as instances of embodied cognition, and, on account of the shared nature of the resulting audiovisual works, as instances of music remediation as well.

Line Rider videos follow the same general rules that determine what is or is not perceived as synchronization between music and image. Donnelly describes a similar process in film music:

Indeed, the relationship of music and image often directly follows the principle of analogy... Sound and image analogues take the form of rhythm, gesture, and movement. For instance, a broad turn of camera movement can often synchronize with a palpable sweep of music.⁸¹

When compared to other media and software, it is however possible to point out some of the specificities of *Line Rider* as a tool for the creation of music animated videos. The open-ended nature of *Line Rider*, compared to rhythm games, as well as the fundamental role played by the already mentioned character, allows for a level of freedom that tends to push away from a strictly notational paradigm, suggesting instead a proximity to the opposite one: dance.

The avatar, known within the *Line Rider* community as Bosh (a shortened version of *Line Rider* creator's name, Boštjan Čadež), represents the single moving object exploring the digital canvas and through this movement tends to embody the trackmaker's subjective interpretation of the musical piece. *Line Rider*'s choreographies are necessarily a much more stylized and minimalistic form of dance, but the fact that its emotional range relies almost entirely on a single character's movement in a virtual space

⁸⁰ Compton, "Casual Creators," 67.

⁸¹ Donnelly, *Occult Aesthetics*, 36.

does point to a similar mode of expression, the main difference being that the emphasis falls on Bosh's overall position in relation to the track lines, rather than the individual gestures of a dancer. *Line Rider*'s main approach to intermedial translation is in fact focused more on the overall movement than codified elements. Despite relying on a computer simulation, it is not built on top of mathematical structures which underly both music and visuals; instead, it interprets the emotional content of the soundtrack. As a consequence, the correlation between visual elements and their musical counterpart is often much less rigid and sometimes unexpected. For example, the character's speed is rarely made to correspond to the overall tempo of the music piece. The synchronization is achieved through highlighting strong beats with sudden changes of direction or interactions between the rider and track lines. Bosh's speed, on the other hand, typically correlates with emotional intensity rather than purely musical parameters. This emphasis on movement resonates with any cinematic production, but it is particularly significant in a software like *Line Rider* which entirely relies on the spatialization of music.

Another peculiarity comes from its status as a physics simulator, however limited in its capabilities. The tension and release dynamic that regulates most musical artifacts is in a way built into the *Line Rider* software itself. Laws that govern momentum and, most importantly, gravity are exploited to become expressive tools: Bosh is often made to fall on the downbeat, capitalizing on deeply ingrained and pseudo-synesthetic analogies that already found their way into language.

Its nature as a physics engine also poses a few questions, namely about its relation to traditional animation. On one hand, the artifacts produced with it are unambiguously categorized as animations, and whether it be considered machinima or not depends on how narrow of a definition is used. On the other hand, it would be extremely disingenuous to consider *Line Rider* an animation software *tout court*. The two fundamental elements—i.e., the character and the track lines—cannot be manipulated independently of each other, and simple black lines are the only drawing tool available. These constitute unnecessary constraints from the perspective of animation software, while they are warranted in the context of a casual creator.

The same limitations also make *Line Rider* videos a very effective and accessible form of intermedial translation. More than traditional animated films, *Line Rider* visualization can support and magnify musical predictability. Centering the entire sequence on a single character focuses the audience's attention, and its nature as a physics engine makes it possible to an-

ticipate their movements on the canvas; this allows the *Line Rider* spectator to “see” downbeats and accents in advance rather than passively following potentially unrelated images in a frame-by-frame animation. Gravity can be, and often is, used to create tension and expectations, which are then resolved by synchronizing the impact with a track line, mimicking the accompanying rhythm.

Conclusion

The main goal of my research has been to explore and further develop the notion of *intermedial translation* as an analytical tool with which to gain new insights into the analysis of audiovisual works, in particular ones that remediate music through synchronized visuals. In order to do so, I began by examining the history and use of the term *intermedial translation*, situating it at the intersection of intermedial studies and translation studies. Therefore, I first employed Elleström’s modality model to clarify the differences entailed by translation between different media compared to verbal-based translation; then critiqued his notion of *media translation*, adopting a functional approach to translation as a way to discuss and refine the analytical tools developed thus far. Finally, I clarified how this perspective can be adapted to audiovisual analysis by engaging with literature on synchronicity/synchronization. Having developed and refined the notion of intermedial translation using sources from intermedial studies, translation studies, and audiovisuality, I put to use these theoretical tools by examining software and practices that result in musically animated videos. In particular, using *Line Rider* as my main case study, I have explored how the specifics of its visuals translate the chosen soundtrack, and how the resulting animated video differs from other forms of animation.

I have argued so far that the musical component plays a fundamental role in shaping the language of many audiovisual works, whether aimed at education, entertainment, or artistic expression, created through casual creators. As a source text in a process of translation, the soundtrack acts as a framework that guides both the creative process as well as the perception of the resulting audiovisual artifact. As I have mentioned earlier, this relation is not unidirectional: whenever an already existing composition is repurposed and displayed as soundtrack in an audiovisual object, the piece itself represents an instance of music remediation. By converting the soundtrack into visuals, each animated video contributes a new interpretation to the

history of the musical artifact it mediates, thus affecting the collective understanding of that particular piece. Together with the many other software and techniques with which audiovisual works are created, *Line Rider* situates itself in the complex network of constant music remediation, which operates through any type of audiovisual production and consumption, either through traditional channels or more extensively mediated by social media.

Further research into how these online communities and practices actually come to be and how they relate to the wider collective media consumption would open up possibilities for a more in-depth analysis than what I have produced so far and seems to be the logical next step. On the other hand, the notion of intermedial translation developed in this article is not medium-specific, and my own application to various categories of software is in fact primarily aimed at demonstrating its versatile nature as an analytical tool. Therefore, its implementation into other fields of studies also represents a way to further test this approach and hopefully achieve new insights through the type of analysis it makes possible.

Works Cited

- Aasen, Solveig. "Crossmodal Aesthetics: How Music and Dance can Match." *The Philosophical Quarterly* 71, no. 2 (2021): 223–40.
- Bailie, Joanna. "Film Theory and Synchronization." *CR: The New Centennial Review* 18, no. 2 (2018): 69–74.
- Bruhn, Siglind. *Musical Ekphrasis: Composers Responding to Poetry and Painting*. Hillsdale: Pendragon Press, 2000.
- . "Penrose, 'Seeing is Believing': Intentionality, Mediation and Comprehension in the Arts." In *Media Borders, Multimodality and Intermediality*, edited by Lars Elleström, 137–49. London: Palgrave Macmillan, 2010.
- Chesterman, Andrew. "Cross-disciplinary Notes for a Study of Rhythm." *Adaptation* 12, no. 3 (2019): 271–83.
- Chion, Michel. *Audio-Vision: Sound on Screen*. Translated by Claudia Gorbman. 2nd ed. New York: Columbia University Press, 2019.
- Compton, Katherine. "Casual Creators: Defining a Genre of Autotelic Creativity Support Systems." PhD diss., University of California Santa Cruz, 2019.
- Cook, Nicholas. *Analysing Musical Multimedia*. Oxford: Clarendon Press, 1998.
- Donnelly, Kevin J. *Occult Aesthetics: Synchronization in Sound Film*. New York: Oxford University Press, 2014.
- . "The triple lock of synchronization," in *The Cambridge Companion to Video Game Music*, edited by Melanie Fritsch, Tim Summers, 94–109. Cambridge: Cambridge University Press, 2021.
- Eco, Umberto. *Experiences in Translation*. Translated by Alastair McEwen. Toronto: University of Toronto Press, 2003.
- . *Mouse or Rat? Translation as Negotiation*. London: Phoenix, 2003.
- Elleström, Lars, ed. *Beyond Media Borders, Volume 1: Intermedial Relations Among Multimodal Media*. London: Palgrave Macmillan, 2021.
- Elleström, Lars, ed. *Media Borders, Multimodality and Intermediality*. London: Palgrave Macmillan, 2010.
- Elleström, Lars. "The Modalities of Media: A Model for Understanding Intermedial Relations." In Elleström, *Media Borders*, 11–48.
- . "The Modalities of Media II: An Expanded Model for Understanding Intermedial Relations." In Elleström, *Beyond Media Borders, Volume 1*, 2–91.
- Englund, Axel. "Intermedial Topography and Metaphorical Interaction." In Elleström, *Media Borders*, 69–80.
- Fahlenbrach, Kathrin. "Aesthetics and Audiovisual Metaphors in Media Perception." *CLCWeb: Comparative Literature and Culture* 7, no. 4 (2005). <https://doi.org/10.7771/1481-4374.1280>.
- Filimowicz, Michael, and Jack Stockholm. "Towards a Phenomenology of the Acoustic Image." *Organised Sound* 15, no. 1 (2010): 5–12.
- Garro, Diego. "From Sonic Art to Visual Music: Divergences, Convergences, Intersections." *Organised Sound* 17, no. 2 (2012): 103–113.
- Giannakopoulou, Vasso. "Introduction: Intersemiotic Translation as Adaptation." *Adaptation* 12, no. 3 (2019): 199–205.
- Jakobson, Roman. "On Linguistic Aspects of Translation." In *On Translation*, edited by Reuben Arthur Brower, 232–39. Cambridge, MA: Harvard University Press, 1959.
- Johnson, Phylis, and Donald Petit. *Machinema: The Art and Practice of Virtual Filmmaking*. Jefferson: McFarland & Company, 2012.

- Kaindl, Klaus. "Multimodality and Translation." In *The Routledge Handbook of Translation Studies*, edited by Carmen Millán, Francesca Bartrina, 257–69. London: Routledge, 2013.
- . "A Theoretical Framework for a Multimodal Conception of Translation." In *Translation and Multimodality. Beyond Words*, edited by Monica Boria, Ángeles Carreres, María Noriega-Sánchez, Marcus Tomalin, 49–70. London: Routledge, 2019.
- Kanellos, Emmanouil. "Visual Trends in Contemporary Visual Music Practice." *Body, Space & Technology* 17, vol. 1 (2018): 22–33. <http://doi.org/10.16995/bst.294>.
- Krumhansl, Carol L., and Diana Lynn Schenck. "Can Dance Reflect the Structural and Expressive Qualities of Music? A Perceptual Experiment on Balanchine's Choreography of Mozart's Divertimento N. 15." *Musicae Scientiae* 1, no. 1 (1997): 63–85.
- Marshall, Sandra K., and Annabel J. Cohen. "Effects of Musical Soundtracks on Attitudes Toward Animated Geometric Figures." *Music Perception* 6, no. 1 (1988): 95–112.
- Miller, Kiri. *Playing Along: Digital Games, YouTube, and Virtual Performance*. New York: Oxford University Press, 2012.
- Milstein, Dana. "Case Study: Anime Music Videos." In *Music, Sound and Multimedia. From the Live to the Virtual*, edited by Jamie Sexton, 29–47. Edinburgh: Edinburgh University Press, 2007.
- Mitchell, Robert W., and Matthew C. Galaher. "Embodying Music: Matching Music and Dance in Memory." *Music Perception* 19, no. 1 (2001): 65–85.
- Montesi, Vanessa. "Translating Paintings into Dance: Marie Chouinard's *The Garden of Earthly Delights* and the Challenges Posed to a Verbal-Based Concept of Translation." *The Journal of Specialised Translation* 35 (2021): 166–85.
- Moseley, Roger. *Keys to Play: Music as a Ludic Medium from Apollo to Nintendo*. Oakland: University of California Press, 2016.
- Naumann, Sandra. "The Expanded Image: On the Musicalization of the Visual Arts in the Twentieth Century." In *Audiovisiology, A Reader*, Vol. 2: Essays, edited by Dieter Daniels, Sandra Naumann, 504–33. Köln: Walther König, 2015.
- Ouzounian, Gascia. *Stereophonica: Sound and Space in Science, Technology, and the Arts*. Cambridge, MA: The MIT Press, 2021.
- Queiroz, João, and Pedro Atã. "Intersemiotic Translation, Cognitive Artefact, and Creativity." *Adaptation* 12, no. 3 (2019): 298–314.
- Rajewski, Irina. "Border Talks: The Problematic Status of Media Borders in the Current Debate About Intermediality." In Elleström, *Media Borders*, 51–69.
- Roux-Girard, Guillaume. "Sound and Videoludic Experience." In *The Oxford Handbook of Interactive Audio*, edited by Karen Collins, Bill Kapralos, Holly Tessler, 131–44. Oxford: Oxford University Press, 2014.
- Schober, Regina. "Translating Sounds: Intermedial Exchanges in Amy Lowell's 'Stravinsky's Three Pieces "Grotesques" for String Quartet'." In Elleström, *Media Borders, Multimodality and Intermediality*, 163–74.
- Summers, Tim. *Understanding Video Game Music*. Cambridge: Cambridge University Press, 2016.
- Tobias, James. *Sync: Stylistics of Hieroglyphic Time*. Philadelphia: Temple University Press, 2010.
- Tymoczko, Maria. *Enlarging Translation, Empowering Translators*. Manchester: St. Jerome Publishing, 2007.
- Woolman, Matt. *Sonic Graphics: Seeing Sound*. London: Thames & Hudson, 2000. Abstract

Abstract

In this article I discuss the notion of intermedial translation in the context of audiovisual works. I begin by presenting Lars Elleström's modality model as it provides the larger framework under which I then specify the scope of intermedial translation. After confronting different interpretations of the concept, I present my main contribution, which links this theoretical notion to the practice of music synchronization. I argue that the resulting approach provides a unique tool for the analysis of audiovisual artefacts. In order to further clarify this analytical perspective and prove its usefulness, I then proceed to apply it to artifacts created through a variety of software, ranging from music visualizers to video games. A look into 2006 Flash game *Line Rider* and the online community that spawned around it completes the article as my main case study.

Michele Rota holds a BA in Philosophy and an MA in Music, Theater, and Film from the University of Milan. Among the main areas of interest are media studies and new media in particular. His research focuses on the relation between animation and digital technologies, often favoring an intermedial approach to these subjects.