# XENO: A NEW PERSPECTIVE OVER AUDIO-VIDEO RELATIONSHIP IN MULTIMEDIA ART

*Enrico Dorigatti*

Sound designer, composer, multimedia artist,
Trento, Italy
**enrico@enricodorigatti.com**

**ABSTRACT**

Aim of this paper is to give a broad overview of Xeno, an audiovisual work realized back in 2019 with the purpose to investigate the relationship occurring between the two main protagonists of the work itself, audio and images, making use of new technologies and employing a multimedia artistic approach. The key of Xeno, indeed, is the particular perspective it offers on the existing correlation between audio and video. Starting with a general explanation of the work with the purpose of introducing the context and matter, this paper will move on to explain the connection to the topic of sonification, providing the key for a contextualized interpretation and focusing on the aforementioned relationship between audio and images. At the same time, the artistic idea and concept at the ground of the whole work will be investigated, taking into account the work and theories of French composers Pierre Schaeffer and Michel Chion. Lastly, a technical section will be proposed to explain the practical approach followed to achieve the final result.

## 1. AN OVERVIEW OF XENO

Xeno is a multimedia work composed of audio and video. Its duration is 3' and 41" and it can be categorized as an experimental short film, in the track of some of the pieces realized by American visual artist Stan Brackhage, such as *Black Ice* (1994) and *Chinese Serie* (2003). With these two early works, Xeno shares respectively the color changes and the style of the visuals showed; with both, it has in common the rapid and unpredictable assembly. From a general standpoint, Xeno can be described as an abstract artwork. However, below, audio and video will be separated and analyzed individually as, although they converge in the same final product, they have different qualities and unique characteristics.

### 1.1. Visuals

Starting from the visual part, it becomes clear that its aesthetic is broadly comprised into the definition of abstract art. Indeed, video is composed of frames whose graphical content does not have a standardized meaning, nor a corresponding one can be found into known objects, animals or shapes. Images appearing during the final part of the work could vaguely recall the physical appearance of microorganisms such as bacteria when observed through a microscope; however, at the same time, the aggressive, shocking colors are far from what these organisms will actually appear like, clearly underlining that we are not engaged in a journey into biology. Far away from the traditional figurative art, according to professor Kendall Walton, we are in the non-figurative one realm: what is considered "abstract" is not "figurative" or "objective" [1], e. g. it does not represent known, real-life objects. Of course, strange colors and alien and unrecognizable shapes are distinctive and characteristic elements which make clearer and sharper this categorization and identification, making it the obvious choice. The visuals have been achieved by colorizing and then manipulating, thanks to specific software, an originally black and white footage. Although the processings applied are quite light and subtle, their contributions, combined together, have originated the colorful result shown.

### 1.2. Audio

On the other side, there is audio, strongly entwined with the visuals. However, it is more difficult to define what is "abstract" in music and, consequently, whether a piece or a soundtrack, as it is the case, should be categorized as that: «It can easily seem that music is naturally, normally abstract, whereas painting is naturally, normally representational.» [1]. Walton says that even abstract visual art is representational, as at least these works lay on the detachment from what is representational, while abstract, "absolute" music stands on its own [1]. Of course, an explanation of the intent and objective, when given directly by the artist, can help to shed light, aiding the classification. But another way could be to categorize the music piece using other of its qualities, focusing on different parameters, such as rhythm, vertical structure, horizontal structure, acoustic qualities. Following this second path and using just few of the many properties possible, we can break-up the structure of the soundtrack of Xeno into two main and well-defined parts. On the one hand, focusing on the nature and type of the sonic objects, we can define the first and main one as a glitch-based one, originating a plot with nuances of experimental electronic music and drone music, with a complexity that follows a *crescendo* trend thanks to the technique of accumulation. Although the definition of glitch itself opens up a quite broad scenario of possibilities and meanings, in general, we can define a glitch as a disruption, malfunction, intentionally caused or not of a physical device as well as of a software tool [2, 3], which eventually leads to the generation of a number of artifacts [3]. On the other hand, the second main section of the soundtrack of Xeno comes in nearly at the end of the work. Changing the properties taken into account and using the vertical structure and the rhythmic pattern as descriptors, we can collocate this second one into the club music category without any hesitation. The reasons behind that choice are clear: regular kick, chasing bass and an overall clear rhythmic pattern make any doubt disappear. If we listen carefully, there is still the presence of a layer of glitches at the ground. However, although it is an element of primary importance to understand the meaning of the whole work, in the context of

this second section it can also assume another meaning, and it can be addressed as a secondary layer with the purpose to add strength and the so called body to the music. As a brief note about the genesis of the sounds that can be heard, everyone of them has been carefully crafted employing creatively custom and fully digital chains of audio effects. Thanks to their adaptability, subtle to extreme processes have been applied to the initial audio material. This material was composed of small pieces of synthesized audio, recorded real-life sounds and samples coming from bent circuits, both taken from consumer devices and custom-built.

### 1.3. A final thought about audio

Concluding this first section, we can say that, although it is easier and more convenient to identify the music present in Xeno taking into account different characteristics and properties compared to the video, it is fair to describe it as abstract, for it has the traits summarized below. While the first point is also a statement of the artist, all of them have been retrieved out of the impressions of people informally interviewed after some listening sessions:

- it is not meant, nor it tries to communicate any message by itself: it is "absolute" music, not "programmatic";
- it is realized by means of electronic instruments and digital processes: the difficulty to understand precisely and without any doubt, during a listening session, what exactly generates a certain sound, emphasizes the perception of impersonality and abstraction;
- coming without any harmonic reference, if we do not take into account some very simple hints unveiled during the last fragment, it is even harder to identify it as a carrier of a message or meaning. Indeed, noise, the sonic category in which glitches are normally aggregated [3], is more often referred to as a source of chaos, acoustic pollution and, more generically, as a waste of energy [4], rather than as an almost infinite fount of sonic and musical objects.

## 2. ARTISTIC MEANING AND CONNECTION TO THE TOPIC OF SONIFICATION

Xeno has been realized following a precise idea; this way, it gained a sharp and well-defined artistic sense and message. However, equally important is its significance within the context of auditory display, as it has a meaning this way too. From both perspectives, what is of primary importance is the tight relationship and interconnection between audio and video; but although the focal point is the same, the two interpretations are different, even if they may be somehow entwined. Following, starting from the artistic meaning and perspective, both of them will be explained.

### 2.1. Artistic meaning

The artistic idea at the ground of Xeno is not purely artistic, as it is mixed with the theories, study and works of Pierre Schaeffer and his former student Michel Chion, and also with the physic behind the production of a sound. The initial idea was to simply realize a soundtrack for the visuals. It should have been a musical piece suitable for the rapid changes of the colors and images, but there was not any other particular necessity or intention yet. Soon after, however, it became clear that more than a mere soundtrack could be realized, as the peculiarity of the video would have allowed a lot of creative solutions and possibilities. The idea that slowly came out was to realize a "living" sonification, something that had to breathe and to be alive together with the images. Eventually, this intention lead to think to the birth of a sound from a more concrete, real-world standpoint. It is well known that objects or substances do have to be excited by a force to start oscillating and producing vibrations, which, under certain circumstances, reach our ears as sounds or noises. It is a combination of elementary physics and acoustics, the kind of basic principles that would be normally discovered during childhood. However, this implies something important and that most of the time is forgotten or undervalued: sound does not exist for itself. As a phenomenon, it is always and only the consequence of a physical action or event, and it can be analyzed, described and studied from a physical standpoint as well as from a psychophysical one [5]; but it can not exist for its own. The soft and relaxing sound of the rain is caused by a myriad of small droplets falling on the ground or on the roof; the gritty and scratchy, exciting sound of a racing car is a mix of the vibrations generated by the pistons moving into the engine, the tires rotating onto the pavement and the pressurized exhausts leaving the muffler. Humans' speeches and animals' calls are generated by means of vibrations too, with air that physically excites different and particular parts of the body. Even an ethereal sound as it is, for examples, the wind, is the result of the physical interaction between huge masses of air clashing with objects, no matter whether they are the leaves in a forest or the façade of a building. Moreover, there is to take notice that is often very easy to associate the sound to its physical generator, or generating event, either because there is a previous knowledge of the sound and its origin or because the object producing it, sharing the same or a near physical place with the listener, is visible and easily spottable. For the first case an example could be a door slamming in another room, while for the second one it can be the aforementioned racing car example. In both these occurrences, the sound is coupled, entwined with its physical source so tightly that it is possible to estimate properties such as the material it is made of, the distance from the listening point, the size of the object; it is even possible to speculate about the type of movement involved in the production of that particular sound. Michel Chion, in his *Guide des Objets Sonore*, pursuing the research undertaken by his maestro Pierre Schaeffer, classifies four types of listening. The first one, originally called *Écouter*, "listening", overlaps perfectly what has been just said, as its description depicts the situation when a sound is a medium, a clue to understand its fount and cause [6]. Digging deeper into this definition, Chion subdivides it further, implementing two sub-categories. The first one, called *Écoute banale*, "banal listening", describes a type of listening in which the listener immediately tries to understand the cause and origin of a sound, and its meaning, without, however, lifting any question about the sound itself as an event and its functioning [6]. In short, not only the imagination and attention of the listener is more often directed to the source of the sound; even worse, the sound itself may not be acknowledged globally, in all its complexity and with all its components and peculiar characteristics, but it may be overlooked and recognized just as the sonic byproduct of an action or interaction between physical objects. Although all of that sheds some light on the power of humans' ears-brain system, from a sound designer's perspective it is often frustrating to know that it does not matter how much care and attention are employed when creating new and complex sounds: even when listening to the most abstract and fancy

ones, a listener will likely try to imagine and virtually realize for them a source and a context. Returning back to the production of Xeno, the objective to design and create a breathing, living sonification was starting to have the profile of a nightmare. Not only there were these problems related to the type of listening practiced by the vast majority of the people, as normally there is to make practice before being able to perform a critical type of listening [6]; there was also the question represented by the abstract visuals, another element which would have distracted the attention of the listener even more carrying it away from the sounds. The decision to realize what is now the soundtrack of Xeno came after a thorough brainstorming, and the final idea was to realize a soundtrack with the precise main purpose, to subvert the conception and hack the perception that sound in general is somehow subordinated to another event. Of course this is how things function in the real world, and this can not be changed; but when it comes to art, one has the possibility to imagine a universe, redesigning from scratch the rules ruling it. Concretely, there was the need to implement a sonification and, more generally, a whole multimedia work which would have prompted the audience to employ the type of listening that Schaeffer and Chion describe as *Écoute réduite*, "reduced listening": a type of listening which takes into account the sound *per se*, decoupling it from any possible source, real or supposed, and meaning it could be carrying [6]. This way, the characteristics of sound are taken into account are only its structure and materiality. Sounds themselves are considered sounds no more: instead, they are described as *objet sonore*, "sonic objects", and each one of them is defined as every sonic phenomenon which is perceived as a coherent whole. In addition, a sonic object is not the material source of the sound nor the physical vibration; it can not be fixed or record, nor notated on a score, and it is not a mood [6]. Reduced listening and sonic objects, are the purest quintessence of sound. But the practical implementation of this conceptions and theories into a music piece, this was the real challenge. The optimal solution turned out to be a sonification which, by strongly coupling together audio and video, makes understanding which one of these two media is cause or consequence of the other difficult. Are visuals the origin of sounds, as it is in films? Or it is the contrary? Or, maybe, there is not a correlation between visual objects and sonic ones, and they share the same space together harmonically, without the ones excluding the others? While working and experimenting on the content of Xeno, some designed elements and strategies turned out to be more effective than others to achieve the desired result, while others were already embedded into the video footage. Some others turned out to be functional only at a later time, after receiving feedback from the listeners. In general, labeled as the most important one, they are:

- abstract material, both visuals and sounds: it is more difficult to assign them a predefined, known meaning, and it makes difficult for the audience to understand whether one object represented in one of the two media is half of a whole, or not;
- concerning visuals, due to the fast assembly rhythm and color changes it is hard to focus the attention on something, as the frame after changes everything. There is not time to think, assimilate what has been just shown, assign a meaning and then move on: what you see, already belongs to the past, it is gone;
- concerning audio, the granular, organic texture that defines the sound for the majority of the duration, enhances the illusion that what is heard is something alive and that it is moving. This is the corresponding effect to the one previously mentioned for the visuals;
- for both visuals and audio, there is a never-ending alternation between, respectively, colored frames and black, or void ones, and between sound and silence. On the one hand, as these events are very short and evanescent, atomic ones, it is difficult to think to them as powerful enough to generate something both visible and audible. However, on the other hand, audio and video are too strongly overlapped in time to result uncorrelated, as both of them start and end perfectly together. What turned out thanks to the informal interviews to the audience after the aforementioned listening sessions, is that the most immediate and easy reaction for the listener was to treat sounds and visuals as separated entities, yet corresponding, "living together" in time.

This last point, the tight correlation in time between sounds and visuals, and black frames and silence, is perhaps the most powerful and effective artifice realized to reach the objective to induce a reduced listening behavior, besides being the most technically interesting challenge. This point is also, among the others, the one more entwined with the topic of the sonification. Indeed, what the last point describes appears to be the strongest effect to keep the audience from coupling instantaneously audio and visuals together in a causality relationship, in which the banal listening would easily categorize every bite of sound simply as a product of a mysterious visual event.

### 2.2. Sonification and auditory display meaning

According to [7], the term auditory display describes «systems that employ sonification for structuring sound and furthermore include the transmission chain leading to audible perceptions and the application context.». Therefore, «Sonification is thereby an integral component within an auditory display system which addresses the actual rendering of sound signals […]» [7]. As Xeno is composed only of data, it does not provide itself a direct a way to listen to sounds, so it is more correct to say that it belongs to the field of sonification, rather than to the auditory display one. A definition to describe what a sonification is, is the one proposed in [8]: «Sonification is defined as the use of nonspeech audio to convey information. More specifically, sonification is the transformation of data relations into perceived relations in an acoustic signal for the purpose of facilitating communication or interpretation.». To further specify, it can be said that Xeno employs a Parameter-Mapping Sonification (PMS), a particular technique by means of which data, mapped to the acoustic attributes of a sound, change them, playing it [7]. Indeed, as mentioned before, the work presents a strong, almost perfectly synchronized relationship between the occurrence of images on the screen and the presence of sound, and vice versa, clearly demonstrating that this is the case of direct mapping, in which the relationship data-sound is one-to-one [9]. It can be said with confidence that the parameter used to drive the sonification is the appearance of colorized pixels on the screen, and that the sonic attribute it controls is the one commonly known as mute, that is the amplitude of an audio track times zero. As aforementioned, the soundtrack of Xeno employs the technique of accumulation: starting from one sound, over time it collects many other ones, ending with a considerable amount of them playing at the same time. Of course, not every sound is the product of a sonification as

Xeno is first of all a multimedia artwork, and the vast majority of them has only an artistic meaning. Actually, only the first sound that can be heard, the one starting at the very beginning, and few other sub-layers appearing from time to time and very similar for what concerns the type of sonic content are controlled employing the PMS technique on the same parameter. It may seems too little, only few sounds among a myriad of other ones, especially as they does not always have a leading role, ceasing the foreground to the others from time to time and especially during the last section. However, even when faint or in the background, overwhelmed by others timbres and sonorities, these sonic, glitch-like textures are still present, carrying on their task and making possible to hear the changes occurring in the visuals.

## 3. PRACTICAL APPROACH TO THE SONIFICATION

### 3.1. Introduction

The practical realization of Xeno, the path followed to realize the final multimedia piece out from the initial idea, has required in particular the development of a custom tool to overcome the technical challenges presented not only by the artistic idea itself, but also by the material available when the work started. Indeed, the real issue was not to colorize the visuals nor to realize the sounds, separately, in the form in which they are now; this has been a straightforward, creative artistic process, one among the others. The challenge was to realize the parameter mapping from the audio to the video. The starting material was the video alone, and before colorizing and manipulating it, it had only two colors: white for the shapes, and black for all the other frames. How to map correctly the content of the video to the audio track, so to achieve the relationship content-sound, void frame-silence? Moreover, the video was a fixed file: if it had been an open project of a video editing and manipulation software, probably the solution would have been straightforward. Deconstructing the problem into sub-tasks, there were two main assignments to carry out:

1. analyze the video file so to detect black frames and colored ones, and changes from a case to the other;
2. convert data to MIDI and send them to the DAW.

It is clearly understandable why at this point a desirable characteristic of the software dedicated to the first task was the embedded MIDI connectivity. As it was the protocol defined to be used to send information to the DAW, the possibility to send this type of messages directly from the detection software would have avoided the need of a third program exactly in-between, with the purpose to convert and pack detection data into MIDI messages and then forwarding them to the DAW. Flexibility and ease of use, of course, were other welcomed characteristics: as the solution was going to be approached with an experimental, trial and error methodology, the risk to need to restart from scratch was consistent. Therefore, the possibility to easily adjust the parameters of the detection and mapping, would have been very useful.

### 3.2. Video analysis and mapping

A thorough research for ready to use software, either commercial or free, highlighted the lack of available, simple enough and specific solutions to carry out the necessary operations. Eventually, the choice has been to build from scratch such a system using the Max/Jitter programming language, for it offers a flexible prototyping environment with plenty of ready to use high-level objects. Most important, it offer the possibility to easily implement MIDI connectivity, and the Jitter modules are designed to handle video files. After the decision to proceed that way and the actual implementation, the resulting system was working as expected. Without digging too deep into the specific behavior of every object used within the patch, it was functioning that way:

1. a video file was loaded into the patch, and played from start to finish at its frame rate, which, in this particular case, was of 60 fps;
2. every frame was sent to an object extracting the four ARGB (alpha, red, green, blue) values from the image;
3. after discarding the alpha one, the other three were tested. As the black was not perfectly black, nor white was perfectly white, a threshold was determined by studying the ARGB values of a number of frames;
4. if all the three values were below that threshold, the frame was reported as black/void and the corresponding mute MIDI value was sent; otherwise, it was reported as with an image, and an unmute message was sent.

As a side note, there was the need to start simultaneously the transport, or playback, of the DAW and the playback of the video, so to have both of them referring to the same frame at the same time. This way, the successive assembly of the two media would have been much more easy. Basically, the solution adopted was to duplicate the start playback message, triggered manually within the patch, one going to the object designated to the video playback, and one, through MIDI, to the DAW.

### 3.3. The DAW

DAW stands for Digital Audio Workstation, a software capable of advanced audio editing and manipulation. There are many choices available, both commercial and free. While almost every DAW is capable of carrying out the most fundamental operations on audio material, such as editing and manipulation tasks, they differ at most for what concerns workflow, flexibility and additional features, such as built-in audio effects or the possibility to load file types different from the common audio ones. The DAW used to realize Xeno was Reaper, as it not only allows to import videos into the timeline, but it also offers plenty of customization options. In particular, it permits the user to define the Custom Actions, which act like macros and can act on almost every aspect of Reaper. A huge list of actions is available to the user, who can use them singularly or in freely defined sets, assigning a key combination or a MIDI command as a trigger. With such a possibility, it was very easy to set up the DAW so to allow the Max/Jitter patch to virtually take control of it. In particular, the actions used were the ones controlling the start of the playback and the mute/unmute automation for track #1 in the virtual mixer. There was no need to define commands for any other track: when there was the necessity to control the mute/unmute automation of another track with the data coming from the patch, it was much more straightforward to rearrange the position of the track itself in the mixer, so to have the first slot always reserved to host the audio targeted to be controlled.

### 3.4. MIDI routing

Thanks to the decision to employ a Max/Jitter patch, there was no need for a third software to convert video data into MIDI ones and send them to the DAW. However, there was the need to virtually connect the two distinct software; to create a pipe for the data to flow in. If working with a hybrid digital and analog setup, it would have been easy to connect a device to the workstation. However, operating systems do not provide native support for the internal routing of MIDI messages from one program to one another. The solution came with loopMIDI, a small yet powerful and free tool which permits the creation and management of virtual MIDI connections. Being minimal and very easy to use, it was simple to create the necessary connection and to set both the Max/Jitter patch and Reaper to use it to respectively write and read data packets.

### 3.5. Overall performance of the whole system

The whole chain demonstrated a very good behavior both in detecting the changes in the content of the video, and in transferring them real time to the mute/unmute automation in Reaper. Once the system had been set up, after some paramenters adjustments it was performing as expected. The possibility to load the video file into the Reaper timeline has greatly simplified and accelerated the process of checking for errors; however, few inconsistencies were found between the visuals and the audio automation, along the whole duration of the work, probably generated by some slowdowns of the machine on which the process was executed. These flaws were corrected manually.

## 4. CONCLUSIONS

This paper has introduced Xeno, a multimedia artwork which employs the PMS to propose a novel approach to audio in the context of multimedia art. The use of a scientific technique for artistic purposes, and the sonification becoming a fundamental part of the artwork, confirms the relationship between sonification and *ars musica* [9]. As a final thought, taking into account the formal definition of sonification and the requisites for a technique to be fully classified as that, proposed in [7], it can be agreed that Xeno employs a sonification technique and that the system built to realize it operates a PMS with direct data mapping.

## 5. REFERENCES

[1] K. L. Walton, “What Is Abstract about the Art of Music?,” *The Journal of Aesthetics and Art Criticism*, vol. 46, no. 3, pp. 351-364, Spring 1988.

[2] H. Davies, *Creative Explorations of the Glitch in Music*, 2004.

[3] R. Menkman, *The glitch moment(um)*, The Netherlands, NL: Institute of Network Cultures, 2011.

[4] P. A. Coates, “The Strange Stillness of the Past: Toward an Environmental History of Sound and Noise,” *Environmental History*, vol. 10, no. 4, pp. 636-665, October 2005.

[5] F. A. Everest, *Manuale di Acustica*, Milano, IT: Ulrico Hoepli Editore, 1996.

[6] M. Chion, *Guide des Objets Sonores*, Prais, FR: Éditions Buchet/Chastel, 1983.

[7] T. Hermann, “Taxonomy and Definitions for Sonification and Auditory Display,” in *Proc. of the 14th Int. Conf. on Auditory Display*, Paris, FR, 2008.

[8] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. H. Flowers, N. Miner and J. Neuhoff, “Sonification Report: Status of the Field and Research Agenda,” Rep., National Science Foundation, 1999.

[9] P. Vickers, B. Hogg, “Sonification abstraite/sonification concrete: An ‘æsthetic perspective ` space’ for classifying auditory displays in the ars musica domain,” in *Proc. Of the 12th Int. Conf on Auditory Display*, London, UK, 2006.