Transferring Visual Patterns to Audio Through a Visuals-driven Algorithmic Approach

Enrico Dorigatti
School of Creative Technologies, University of Portsmouth
Eldon Building, Southsea, PO1 2DJ
Portsmouth, UK
enrico@enricodorigatti.com

ABSTRACT

Xeno is is an audiovisual artwork characterised by a tight connection between audio and visuals. The artwork aims to foster speculative reasoning on the nature of sound and challenge its nature within the audiovisual context, where it is usually seen as a product of physical actions.

The cohesion between the visuals and the audio required the latter to mirror through a sequence of silences and sounds the pattern of black frames and images exhibited by the former. However, as a manual approach to this task was impractical, the algorithmic system described in this paper was specifically developed.

Author Keywords

Experimental art, Algorithmic composition, Integrated audiovisual composition

CCS Concepts

•**Applied computing** → **Media arts; Sound and music computing;** •**Software and its engineering** → *Domain specific languages; Visual languages;*

1. INTRODUCTION

Xeno [5] is an audiovisual artwork fitting the abstract aeasthetic [1] and characterised by the strong linkage between its visual and sonic components. The aim of the artwork is to challenge the notion that sound is always a product a happening. Sound is a vibration produced by any physical action, such as two objects colliding, travelling through a medium (usually the air) and reaching the eardrums [6].

Within the audiovisual context, the nature of sound as a product of a different action is frequently explicit. In movies, for instance, sounds are often *diegetic*, thus realistic and usually studio-based reconstructions of sonic events related to the action visually represented [8] and aimed at increasing the sense of immersiveness, engagement, and reality of the movie.

Through the audiovisual medium, *Xeno* aims to challenge and reconsider this action-reaction (causal) linkage. This purpose recalls the concept of *reduced listening* proposed by Pierre Schaeffer, suggesting to approach listening by considering only the sonic matter and its properties rather than the context surrounding the sonic event [3].

In *Xeno*, the abstract aesthetic of the artwork, hardly recalling extra-musical and extra-artistic references and freeing the audience from the dominance of reality [1], jointly with the rapidly sequenced discrete sonic and visual events, contributes towards straining the certainty that sound is a consequence of the visual events shown, opening to speculative reasoning and new interpretations. Understanding which of the two mediums, visuals or audio, is cause or consequence becomes difficult to determine, to the extent that the existence of a causal relationship becomes blurred. Notably, the Greek word giving name to the artwork indicates something different, extraneous.

A notable aspect of *Xeno* is that the visuals and some audio layers within the soundtrack tightly share the same pace and sequence, a characteristic contributing to blurring the cause-effect relationship between the two mediums. To achieve this feature, in place of a manual (mouse-based) approach, an algorithmic system was realised to achieve a faster and more efficient workflow and a more precise result.

1.1 Visuals

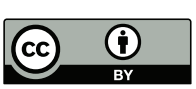
The visuals of *Xeno* are fast-paced sequences of abstract images interleaved with black frames. Such images are limited by a circle in the middle of the screen. Outside that circle, the frame is pitch-black. Throughout the artwork, the figures and colours evolve, and abstract shapes slowly replace the granular textures appearing at the beginning.

The visuals presented by the artwork derive from an originally black-and-white (BW) short abstract movie, subsequently manipulated through saturation and displacement effects. The processes enacted added the colours and doubled and shifted the images. "Abstract art by definition does not depict recognizable objects, but instead uses shapes and colours for their own expressive value in a nonrepresentational way." [2, p. 256], and this definition perfectly aligns with the visual content of *Xeno*. Three stills from the artwork are showcased in Figure 1.

1.2 Audio

Concerning the sonic elements of the soundtrack, they were obtained by manipulating sound samples of different origins, including field recordings and circuit-bent devices [7]. Diverse chains of digital effects were utilised to achieve various sonic elaborations of the pool of original samples, and the results were employed as sonic layers [4].

The composition of the soundtrack roughly followed the evolution of the visuals and employed the *accumulation* tech-



2nd Audiovisual Symposium 2024, December 6, 2024, Dalarna Audiovisual Academy (DAVA), Dalarna University, Falun, Sweden.

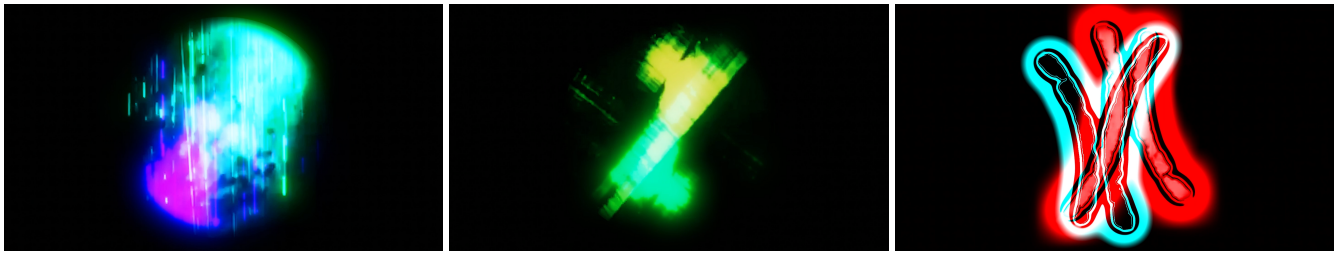


Figure 1: Three stills from *Xeno*. It is possible to appreciate the granular visuals (1st still) slowly morphing into the abstract figures of the final part of the piece (3rd still). Stills 1 and 2 are limited in a circular shape.

nique. Therefore, whilst a noisy layer recalling granular synthesis [11] can be heard from the beginning of the artwork and until the end, other sonic elements join and merge at a later time, creating a glitch-based [10] sound bed.

In addition to more experimental sonic layers, the last part of *Xeno* recalls trance music and relies on a bass synthesiser, a melodic layer, and a drum pattern. However, it is still possible to spot the experimental sonic layers lurking below this section of the artwork.

2. AUDIOVISUAL SYNCHRONISATION

In the making of *Xeno*, the soundtrack was realised on top of the pre-effected visuals. As the purpose was to strengthen the connection between the two mediums whilst blurring any causal relation, the most prominent audio layer was designed to follow the pattern showcased by the visuals, thus being audible only in the case of non-black frames.

REAPER was chosen as the Digital Audio Workstation (DAW) for making the soundtrack. The main audio layer was realised as a granular-like sequence through heavy sound processing based on the free Glitchmachines plugins.

For the sounds to follow the pattern of the visuals, this latter file was loaded into REAPER to be scrolled frame by frame and manually draw the mute automation of the audio track accordingly, thus muting it in the case of black frames and unmuting it otherwise. The mute parameter was chosen for its switch-like behaviour (on/off), more useful in this context if compared to a multi-step parameter like volume, as shown in Figure 2.

However, the difficulty of such an approach became soon clear for two main reasons. On the one hand, the high number of visual transitions and the mouse-based approach imposed a slow workflow. On the other, there is to consider the intrinsic imprecision of such an approach, often making it necessary to revise the shape of the automation not perfectly aligning with the visuals.

2.1 Visual detection and automation lane

An algorithmic system alternative to manually setting the mute automation for the audio track was realised through a Max/Jitter patch. Due to the specificity of the task and the REAPER-based workflow, custom-building a system was preferred over employing any possible ready-to-use, either commercial or open-source software available. The system was developed in the multimedia-oriented visual programming language Max/Jitter, as this environment offers a sandbox for fast prototyping through a large set of ready-made high-level objects connectable in functional networks[1]. The system realised was organised in a patch comprising four sections, showcased in Figure 3.

Video management (green)
The task of the rightmost section of the patch is to load and play the video file within the Jitter environment. Its main components are a video playback object and other functional blocks gaining data from the loaded media file and its frames as they are processed.

Image detection (blue)
The per-frame data retrieved are sent to the leftmost section of the patch, dedicated to analysis. This section is the core of the algorithm, built around the *jit.3m* object and a set of logic conditions. This section decomposes each frame in a four-value tuple containing the discrete RGBA (Red, Green, Blue, and Alpha) mean value of each frame. The alpha value is discarded, whilst the other three are tested against a threshold t. If all the tests report that the value is below or equal to the threshold, the frame is considered black, otherwise it is considered as showing an image[2]. A MIDI value related to the result of the test is outputted to be sent to REAPER.

Playback control (red)
The topmost section is responsible for setting the MIDI channels and starting the playback of the video file from the first frame.

MIDI out (yellow)
The last section, at the bottom of the patch, is responsible for sending to REAPER the MIDI value resulting from the test.

2.2 Max/Jitter and REAPER

Within the professional audio community, REAPER is often praised for its remarkable flexibility, allowing to configure the software and its behaviour around their needs instead of being forced to a specific workflow. Within the tools available to the end user and allowing for in-depth customisation of the DAW, custom actions, as the name suggests, trigger actions (operations) at certain conditions.

In the realisation of the system described, custom actions were set for the first (topmost) track within the project, which in turn was set to listen for incoming MIDI messages

[1]Max and Jitter are part of the same software distribution. Yet, whilst Max objects are general-purpose and MIDI (Musical Instrument Digital Interface)-oriented [9], Jitter ones are specifically designed to operate on images and video streams. The distribution also comprises the MSP objects specifically designed to operate on audio streams. Each object type employs patch chords with different colour coding. Plain Max objects employ solid grey connections, MSP ones use back and yellow chords, whilst the Jitter environment utilises green and black connections.

[2]The value of the threshold was calculated as the average value of different randomly selected coloured frames.

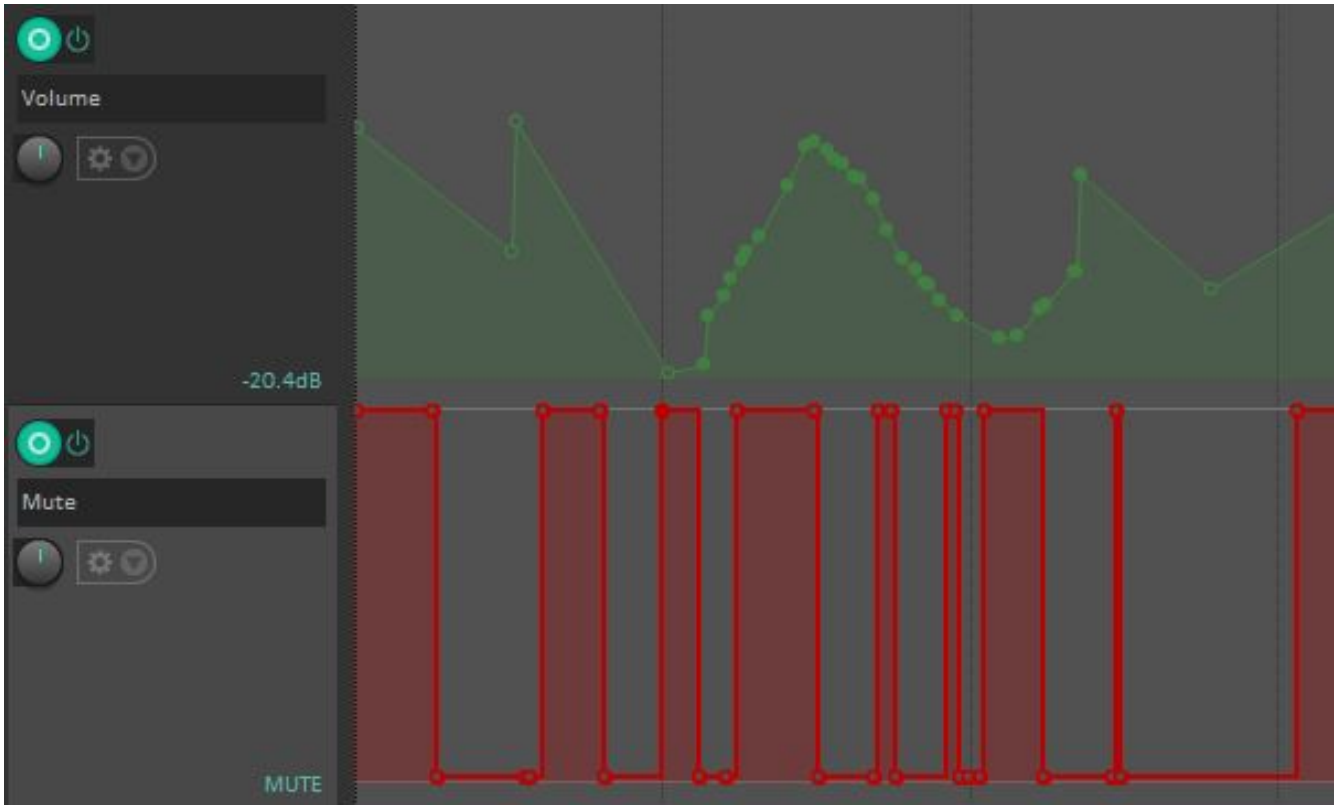


Figure 2: Automation lanes on an audio track in REAPER. The green lane controls the volume of the track, whilst the red automation, exhibiting a on/off behaviour, controls the mute parameter.

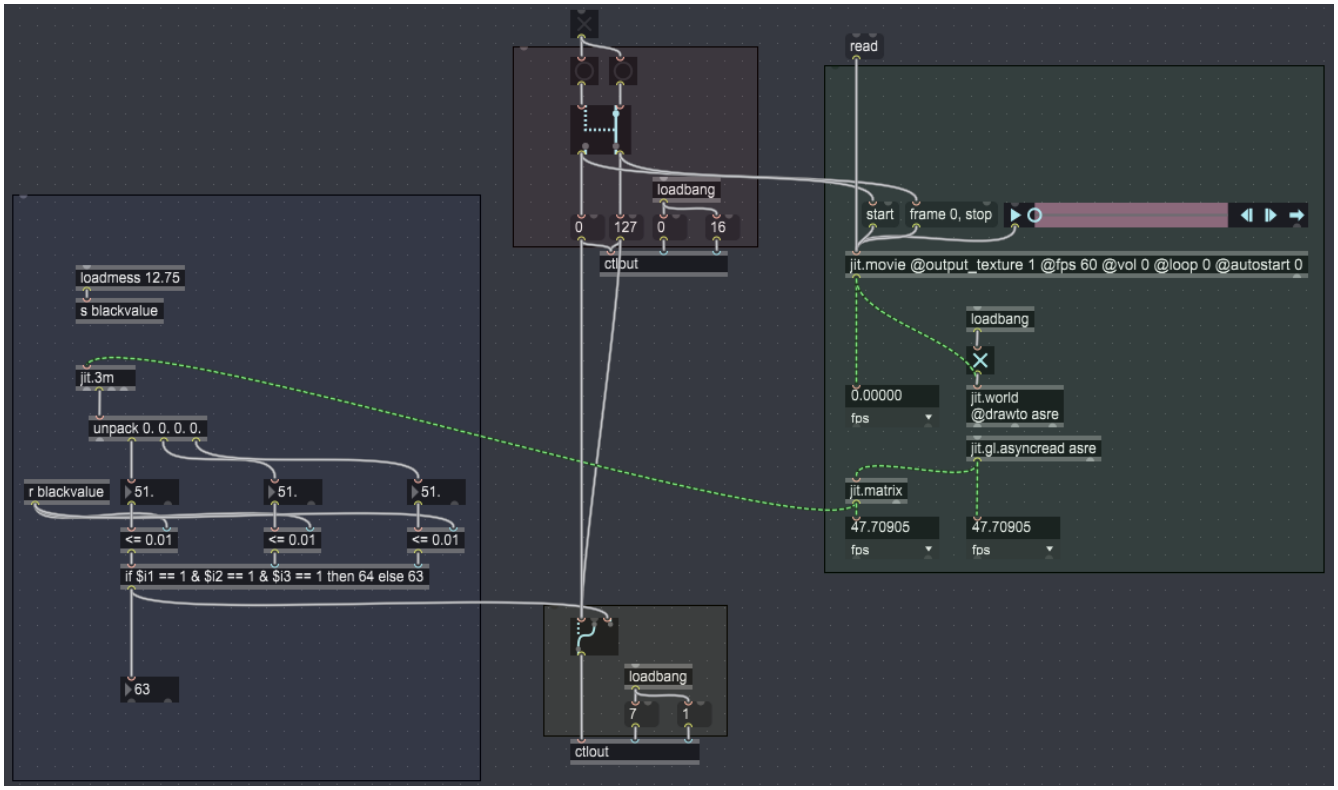


Figure 3: A screenshot of the Max/Jitter patch described in Subsection 2.1. The four main blocks it is divided into are highlighted through a coloured background.

from the channel determined in the Max/Jitter patch. Custom actions were subsequently set to mute or unmute the track according to the incoming MIDI message. This process leveraged the mute automation, set in *write* mode.

As the process employed two software, synchronisation became an issue. Writing audio automation consistently with the visuals (i.e. setting an automation point in REAPER at the exact position in time of the frame analysed in Jitter) required starting the playback (transport) within the two software simultaneously. Another custom action was set in REAPER to start the playback from the beginning of the timeline at the reception of a specific MIDI message sent at the start of the playback in Jitter. In Jitter, a similar mechanism was implemented to set the playhead position at the initial frame before starting the playback.

3. CONCLUSIONS

The system proved worthwhile and exceeded expectations concerning precision and accuracy, with the automation line showing consistency with the pattern of the visuals. Some minor tweaks were necessary as some black frames presented noise and were erroneously interpreted as holding images. However, these few interventions required a negligible amount of time.

After writing the automation lane for the first and main layer of the soundtrack outlined in Section 2 with the system described, other audio layers underwent the same procedure. Yet, their presence is not as noticeable as the main glitch-based, granular-like track, nor are they present for the entire duration of *Xeno*.

The system described is a creative and effective homebrew solution addressing a specific need. While designed for a specific use case, it can be adapted to some extent for other scenarios. For instance, the threshold could be adjusted, and different REAPER custom actions could be triggered by individual RGB values.

References

[1] V. Aviv. What does the brain tell us about abstract art? *Frontiers in Human Neuroscience*, 8:85, 2014.

[2] H. Brinkmann, L. Commare, H. Leder, and R. Rosenberg. Abstract art as a universal language? *Leonardo*, 47(3):256–257, 2014.

[3] M. Chion. *Guide to sound objects*. Éditions Buchet/Chastel, Paris, 2009. (J. Dack and C. North, Trans.). Original work published 1983.

[4] E. Dorigatti. A new perspective over audio-video relationship in multimedia art: A case-study. In A. Andreopoulou, M. Droumeva, J. W. Newbold, K. McMullen, and P. Vickers, editors, *Proceedings of the 26th International Conference on Auditory Display (ICAD 2021)*, pages 277–281. The International Community for Auditory Display, June 2021.

[5] E. Dorigatti. Xeno. *Sonic Scope*, 2023(1), June 2023.

[6] F. A. Everest. *Manuale di acustica: concetti fondamentali, acustica degli interni*. Ulrico Hoepli, Milano, 1996. (G. Bertinotti and R. Minerva, Trans.).

[7] Q. R. Ghazala. The folk music of chance electronics: Circuit-bending the modern coconut. *Leonardo Music Journal*, 14:97–104, 2004.

[8] M. Huvenne. *The Audiovisual Chord*. Palgrave Macmillan, 2022.

[9] L. A. Ludovico. *MIDI: Una guida completa al protocollo, alle estensioni e alla programmazione*. Milano University Press, Milano, Italy, 2021.

[10] R. Menkman. *The glitch moment (um)*, volume 4 of *Network Notebook*. Institute of Network Cultures, Amsterdam, 2011.

[11] C. Roads. *Composing electronic music: a new aesthetic*. Oxford University Press, New York, USA, 2015.

APPENDIX

A. SCREENING LINK

The full version of *Xeno* is available from the online artistic programme of CMMR 2021 - 15th International Symposium on Computer Music Multidisciplinary Research. The artwork is accessible from the link provided below.

Warning: *Xeno* presents flashing images in rapid sequences, which might trigger photosensitive individuals.

Xeno: `https://www.youtube.com/watch?v=BSqugOSoo6I` (last access: November 17, 2024).