Apollo e Marsia: Time and Memory under Tension

xCoAx 2025
13th Conference on Computation, Communication, Aesthetics & X
Dundee, UK
2025.xCoAx.org

This paper describes a hybrid installation - *Apollo e Marsia*, based on a painting by Tintoretto - incorporating music performance on video, physical instruments and machine listening. It explores the dynamics of musical memory, particularly directly after performance and under stress. Through the use of physical instruments and machine listening, it proposes to situate the listener/viewer directly in the environment of the work. Video is shown to be an integral part of contemporary engagement with art music. The work serves as an example to reexamine the status of music as a temporal object and the nature of Stiegler's "tertiary retention", and suggests an alternative way to engage with the arts of time.

1. Source

In 1545, the satirist and blackmailer Pietro Aretino received two paintings from Tintoretto, for the ceiling of his bedroom in Venice. One of these represents a story from Ovid: the musical contest between the god Apollo and the satyr Marsyas, who had challenged the god to a contest ("*The Contest between Apollo and Marsyas*," Wadsworth Atheneum, Hartford Connecticut) (Fig. 1). From Ovid we know the gruesome end that is to befall Marsyas: he must inevitably lose the contest and will be flayed alive (Ovid 1986). But Tintoretto depicts the moment when they have both just finished playing, before judgement is passed; Apollo holds a viola da braccio with a clearly visible sympathetic string, Marsyas a long woodwind instrument. In mid-gesture, they both appear to have just played their last note. Midas is consulting with his fellow jurors, accompanied by a glamorous female assistant (game show formats have a long history). Painting – this painting – does not work like a photograph; there is no punctum here. Time is both compressed and expanded; the painting holds us suspended, dizzy, between its immediate past and its immediate future.

For the two competitors, this is above all a moment of memory – an expanded, stretched moment, because they cannot know when the judgement will arrive. Except in the most unengaged of circumstances, musical time is not experienced in minutes and seconds. In the general case (favourite karaoke songs excluded) the memory of

Jonathan Impett
Orpheus Instituut, Ghent, Belgium
jonathan.impett@orpheusinstituut.be

Keywords installation, memory, sound & video, music, machine listening

DOI 10.34626/2025_xcoax_005

musical experience is not the replaying of a soundfile from start to finish. Intermodulating patterns of form, salience or reference are further complicated by the differential dynamics of the various cognitive mechanisms that constitute musical memory. The resonance of stress makes the immediate memory of personal performance yet more complex, and in the highly-charged moment depicted here, still more so. Each musician is trying to remember, to assess their own performance. In addition, each is also trying to recall that of the other, and to reckon their relative advantages or weaknesses. The moment depicted is of a particularly binary nature (live/die), but the general situation is familiar to any performer. The painting is thus highly suggestive of an intense behaviour of musical memory, in which two sets of actions will inform each other increasingly with their constant re-imagining. It also presents a situation, an emotional state, acutely resonant to any experienced performer.

Fig. 1. *The Contest of Apollo and Marsyas (detail), Wadsworth Atheneum, Hartford, Connecticut*

2. Work

In its complete form, *Apollo e Marsia* is a hybrid installation consisting of two 85-inch UHD LED screens, positioned opposite each other at a distance of 4 meters, four computers, two sound systems and two physical instruments. One screen displays a filmed performance by a musician playing a viola d'amore (a seven-stringed bowed instrument with seven further sympathetic strings below the bridge), the other playing an alto flute - performances of roughly 23 and 25 minutes respectively (Figures 2 and 3). Each screen is surrounded by a pair of loudspeakers mounted at ear level, a microphone above the screen and a physical instrument. The sound is played in 5.1 format. The ear-level speakers present the instrumental sound untreated. Two channels of processed instrumental sound are played through the physical instruments: that of the viola d'amore through a pair of long transparent acrylic tubes, that of the alto flute through an instrument, likewise transparent, with two 3-meter strings. The processed memory of each performer is thus acoustically modulated by the means of sonic production of the other. Sound from the tubes is transmitted through a loudspeaker built into the base of each. The strings are excited by audio signals transmitted through the *body* of two 48v motors. A mac mini behind each instrument handles the video and soundfiles. A second computer on each side runs a k-nearest neighbour engine, listening to the whole through the microphone, responding on the basis of its impressions accumulated through each 24-hour period, and sending its output through the physical instru-

ments. The fifth audio channel of each 5.1 file serves as a constantly varying control to this simple AI, determining the level of its listening.

The listener/viewer is situated between the two protagonists. Their choice of orientation is probably most influenced by the degree of activity in either performance, or perhaps the strangeness, the unattributability of certain sounds from the physical instruments (Fig. 4). However closely their listening is focused on a single source, they will hear all the strands of sound simultaneously: the treated and untreated sounds of recorded performances, and the constant dialogues between the two musicians and between present and past. Duration of attention is clearly beyond the artist's control in an exhibition context, and as such too often ignored in sound art. Here the structure of the two recorded instrumental compositions – each consisting effectively of four or five 3' - 8' movements separated by sonic interludes to still images of the instrument alone – provides the clearest background temporal architecture, but in a constantly changing phase relationship.

3. Challenge

The work seeks to explore under temporal magnification what in experience would be a highly time-compressed, stressful, evaluative memory behaviour. A musician's mind would be scanning their memory in a highly nonlinear and unconsciously-directed manner, jumping between moments of positive and negative salience from both performers, quasi-arbitrarily fixating on certain phenomena, exaggerating or ignoring others, and the whole time attempting to construct some informal, n-dimensional score.

At the same time, *Apollo e Marsia* searches for an alternative path to any inherited live-recorded binary in the apprehension of performed music. Video, we should note, is an increasingly common way to engage with art music performance, to the extent that a) musicians feel obliged to add a video component to their output, however static, live or online, to encourage engagement, and b) audiences will studiously watch almost any video feed in a live situation, such that it becomes a conventional skeletal throughline for much experimental music. Stiegler points to the functional role of such passive engagement in the case of cinema (Stiegler 2010, 10). Rather, video might better be considered not as a substitute or a support, but rather as a vital entry point for engagement with music. Banal as this observation may seem, it has significant implications for how we treat such material. The video performance ceases to function primarily by reference; there is no time or place other than the present of the viewer, except for the mythical.

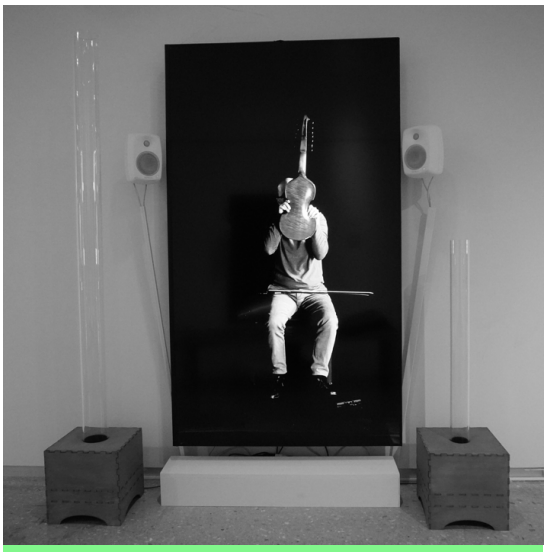

Fig. 2. *Apollo (Marco Fusi, viola d'amore)*

Fig. 3. *Marsia (Richard Craig, alto flute)*

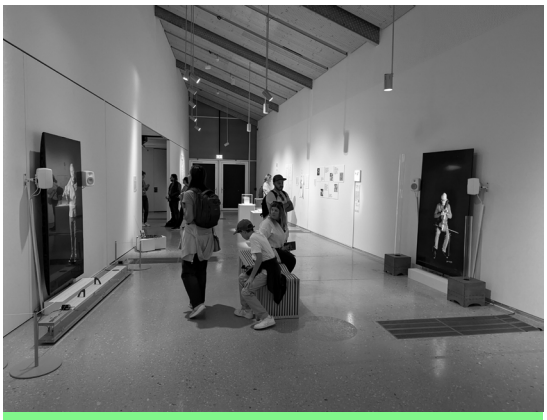

Fig. 4. *Apollo e Marsia in situ (Exhibition Musica ex Machina, EPFL Pavilions, Lausanne, September 2024)*

In the absence of familiar musical form (song, dance, or the norms of the nineteenth century, for example), video affords a thread of linearity that frees the extraordinary range, modality and polyphony of real-time musical memory to do its work. Filmed performance is thus central to the attempt of *Apollo e Marsia* to explore such phenomena, and the ways in which they interact with the ambiguities of liveness and presence.

4. Representation and the dynamics of memory

The viewer/listener is confronted with multiple layers of work: 'work' both in the nominative sense – as constructs with their own mode of production, form and identity – and as layers of different kinds of manipulation. In *Apollo e Marsia* as experienced these are folded together inextricably – from Ovid and Tintoretto to the AI-mediated sounds of the exhibition and its viewers – and yet patterns and resonances across these layers emerge from their interaction.

At the heart of each of the two filmed compositions lies historical material – material to which the filmed performers themselves only have indirect access, through the mediated, transformed, compressed and stressed memory of their own elaborations of sources which were never stated directly in the moment of performance. Two hymns to Apollo inscribed on stone were found at a temple in Delphi: *Paean of Athenaios* and *Paean of Limenios* (Pöhlman and West 2001, 62-85). The pitch indications accompanying the texts are widely regarded as the earliest instances of Western music notation. They themselves can be seen as records, as encodings of hymns sung at games, rather than prescriptive instructions in the sense of a modern score. The surface itself is distressed and fragmented, such that marks are lost. And the process of decoding, beginning with Reinach (1893), still leaves ambiguities – artefacts of cultural memory.

Rather than decoding, the successive re-inscriptions and re-foldings of this material might be seen as moments of transduction. The source material and its various subsequent manifestations do not embody or transmit any kind of message. They are treated as patterns of energy, revealing new aspects of their internal dynamics as they are brought into spaces of increasing dimensionality. The first such transformation happens in the symbolic space of MIDI representation of the ancient source material, then becoming LISP data for manipulation in the computer-assisted composition environment *OpenMusic* (https://forum.ircam.fr/projects/detail/openmusic/), the initial stage in the production of scores for the instrumentalists. Time, sound, memory and movement are all represented as curves

in this environment, rescaled and shifted to produce micro- and macro-wave phenomena as they modulate each other. Wave behaviour and phenomena emerged as central structuring paradigms in working across the multiple modes of representation required by the different environments of generation, manipulation, processing and construction (Viola 2013). Embodied memory plays a part, as the hand and finger movements of each player and the string-crossing patterns of the violist are encoded in the same context.

The relationship between layers of memory (echoic, short- and long-term) and musical form has been studied in detail (Snyder 2016), and fundamental play of memory and anticipation in musical listening is well established (Huron 2008). Such studies relate largely to sequential attention to music that is 'well-formed', not to the multi-directional, fragmented surface of post-performance recall. They do however point to the inextricability of memory and forward-directed imagination. Indeed, research suggests that these are essentially the same mechanism (Budson et al. 2022). Under normal conditions, the principal components of recall of sequential information are well rehearsed: recency, primacy, contiguity, associative asymmetry and semantic proximity (Kahana and Miller 2013). These are in any event nonlinear, subject to progressive unbalancing of focus and increasingly informed by external associations. In situations of stress, such phenomena are strongly further modulated by feedback loops – loops that can themselves proliferate – leading to new artefacts that play no part in the 'original' experience (Tomé et al. 2024). The likelihood of such artefacts diminishes with temporal distance from the stress-inflected moment; more distant recall is clearer about distinction, context and sequence.

In this state of non-intentional self-evaluation, obsessive re-rehearsing of passages is modulated by a polyphonic recall of the performances of both musicians, on time-scales ranging from the complete performance to individual gestures. This material, encoded as waveforms, produces both progressive transformation and sudden switches of state. Passages of sonic and physical memory dissociate and recombine; emergent artefacts become part of the aggregate resource. There is a continuity between this human-directed compositional process and the machine learning by which it will be further mediated; in both cases, a searching for coherence and development is continuously redirected by the emergence of a rhyme or analogy. Machine listening becomes a mode of compression that in its overview and arbitrary access allows the generation of coherent but surprising new material.

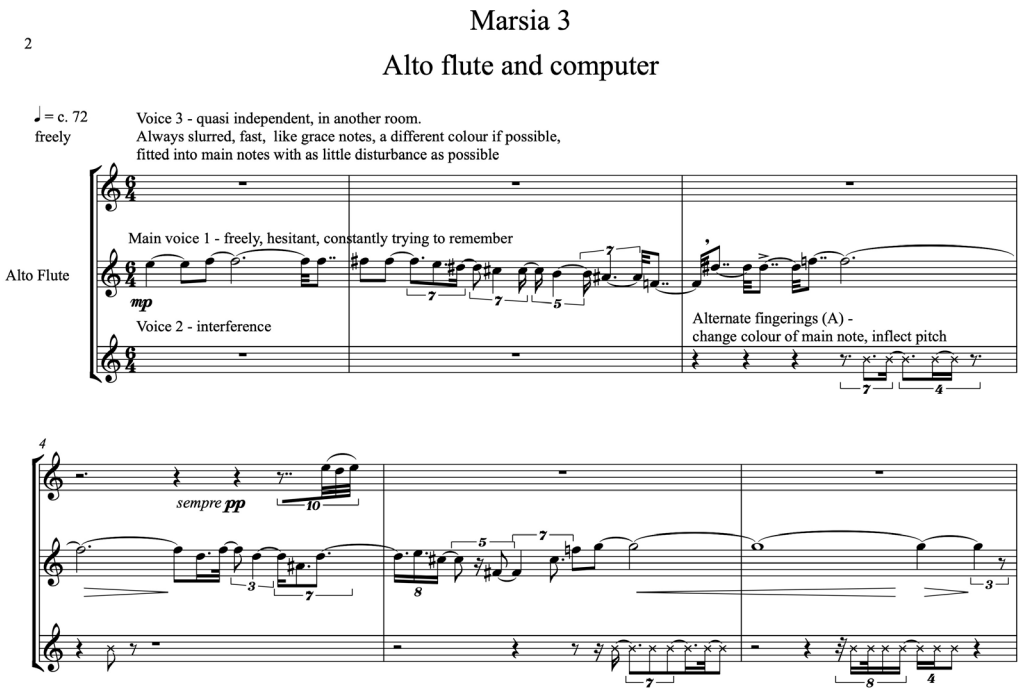

Fig. 5. Marsia, section 3 of score.

In a typical passage, hand movements from one area of performance drive the redistribution of certain pitch elements from another, modulating the shape of an obsessively recalled phrase that links associatively across the source material, the amplitude or resonance envelope of the whole shaped by an insistent central line and interrupted by passing, ungraspable attempts to take the attention elsewhere, to an emerging, unformed memory more remote still. Analogous processes inform the editing of the material this generated as it is re-encoded as Music XML to be manipulated anew in Sibelius notation software. Sometimes this polyphony is assimilated into a single line for the performer; elsewhere the assimilation of these elements in performance itself becomes another layer of 'work'. In some instances, they are presented with multiple staves of notation representing different modes of memory – degrees of certainty, intrusiveness or inaccessibility (Fig. 5).

The recorded performances are then subject to analogous processes of folding, filtering, nonlinear recall and emergent false memory to produce the audio tracks played through the physical instrument accompanying each screen. These are implemented in *Max* (https://cycling74.com), now using a further level of data derived from audio analysis in combination with the symbolic and wave-form material from the initial stage. These polyphonic memory traces are filtered and heard through the means of sound production of the opponent (flute through strings, viola through tubes), while their own sound emerges directly through loudspeakers.

5. Physicality

Each player's memory of their own performance – their sound, invention or virtuosity – is mediated by their attempt to remember and evaluate the performance of the other. While the direct ('live') instrumental sounds are heard from speakers either side of each screen, the sonic memories of player – both Max-processed instrumental sound and AI generated output – are modulated through a physical instrument that represents their opponent. In both cases, sound emerges from a void, a transparent empty space.

Those of the alto flute are heard through two braided steel strings each 3 metres long, mounted on a rectangular cuboid soundbox with transparent acrylic sides, a massive wooden base and a 3mm Alaskan spruce soundboard (kindly supplied by our industrial partner, Chris Maene Pianos). A central laser-cut rose is copied from that of an Egyptian *Ud*. The strings are connected at each end with steel turnbuckles, allowing tuning under tension, and to the soundboard by means of asymmetrical bridges. Modelled after the bridge of the mediaeval tromba marina, the leg not under direct pressure from the string is free to drum against the soundboard at resonant frequencies. After much experimentation with exciters (loudspeaker drivers redesigned to operate in direct context with resonant bodies) it became clear that driving the strings in this way to an appropriate sound level led to unacceptable distortion. Following further work with motors, with different textures on the shaft, it transpired that the motor body itself provided the cleanest transmission across the frequency range. Each string is driven by a single motor, which receives one of the processed 5.1 audio channels. The tuning of the strings derived from *Paean of Limenios*, the flautist's source material.

The analogous sounds of the viola d'amore are heard through two clear acrylic tubes, mounted vertically above loudspeakers – a sort of memory organ. Their lengths – 1940 and 1040 mm respectively – reflect the pitch centres of the Apollonian hymn which forms the basis of the viola's music, *Paean of Athenaois*. The tube diameters follow the practice of *Normalmensur* organ pipe scaling, to encourage appropriate resonance (Mainstone 1998).

6. Obsessive listening

The shifting presence of listeners becomes part of their own experience. The microphone above each screen feeds a second machine. Each side of the installation listens to the room from its own perspective, hearing its own instrument-processed output, that of its opponent, and movement and conversation in the exhibition space.

On each side, a Python machine listening engine then responds, exhibiting a similar set of memory phenomena but now brought into the common present. Incoming audio is sliced using flucoma's 'novelty' analysis (https://www.flucoma.org), and slices are processed by a k-nearest neighbour engine to build a dynamical model of mfcc (spectral) and pitch dimensions. Incoming audio is also matched to the accumulated latent space, according to variable values of temporal distance, probability of prediction, and continuity. The audio is handled using the *librosa* and *pyaudio* libraries. The listening sensitivity of each of the AIs varies continuously and is controlled by the fifth audio channel of the respective performance video. The systems reboot every twenty-four hours, to avoid a terminal accumulation of data, but within that time period the two musicians are listening to themselves, each other and the surrounding environment. The two aural perspectives create two different memories of events, each with their own dynamics and internal associations. The difference in duration between the two video loops avoids precise coincidences and prediction confirmation. Each engine outputs its responses through its respective physical instrument; the constantly reforming memory of each musician is filtered through the sound of the other – haunted by its ghost.

7. Temporal object or constellation?

In developing his notion of tertiary retention, Bernard Stiegler discusses cinematic time, to observe a fundamental relationship between the time of music and that of the moving image; he expends more words on music than on image. His starting point is Husserl's notion of a "temporal object", which Stiegler describes thus:

> [...] a properly temporal object is not simply "in time": it is formed temporally, woven in threads of time – as what appears in passing, what happens, what manifests itself in disappearing, as flux disappearing even as it appears. And the properly temporal object is the ideal object constituting the temporal fabric of the stream of consciousness itself, since the flux of the temporal object precisely coincides with the stream of consciousness of which it is the object. To account for the structure of the temporal object's flux is to account for the structure of the stream of consciousness of which it is the object. (Stiegler 2010, 14)

Husserl's paradigmatic case is melody. In his exegesis, Stiegler expands on that case to propose an interesting distinction between music and sound or noise. He focusses on the note – an apparently

modest concept of extraordinary plasticity: a sound, a frequency, a representation, an instruction. The role of melody here opens up important questions regarding the nature of tertiary retention and its relationship with the temporal object. Stiegler appears to continue Husserl's implicit assumption that a melody is a well-formed construct, amenable to apprehension in its entirety as a unitary object. This may be the case for the simplest song-forms – indeed, that is their strength. But the story of historical western art music could be seen precisely as a search for ways to play with, expand or subvert that possibility. In this respect, Stiegler's tertiary retention may be culturally valid but personally insufficient – that is, art music uses the distancing of inscription and mediation to afford modes of retention beyond the tertiary. Alva Noë's dance-choreography distinction has resonance here art as a human organisational and re-organisational practice – except that music has a deep technicity that right down to its simplest forms engages the most affective strands of human non-referential sense-making (Noë 2015, 13-18).

As common norms of reception have dissolved, modes of engagement with temporal objects have evolved; we begin to study our individual response. *Apollo e Marsia* is an attempt to acknowledge this in the construction and presentation of the work itself, to fold inscription and retrieval, compression and generation into a continuous process. The emergent surface, however, rather approaches a shifting constellation of Deleuze's cinematic "crystals of time" in their third state, a constellation the origins of which are known only through their traces:

> The crystal caught in its formation and growth, related to the 'seeds' which make it up. In fact there is never a completed crystal; each crystal is infinite by right, in the process of being made, and is made with a seed which incorporates the environment and forces it to crystallize. (Deleuze 2013, 92)

8. Conclusions

As they change their own stance and gaze relative to the image, the viewer/listener can engage in slower, more intimate study of the performer, of their features and movement, than is ever possible in live performance. The attempt here is to draw the listener in to search for figure, to dynamically infer process, to build their own space of potential. As we reflect on the condition of music in general, from the perspective of our own technologically-informed moment, perhaps this reminds us that the musical work has always been a virtual entity,

never fixed; in its acoustic surface there is no original. We co-reconstruct and co-imagine in our personal version of a present. Likewise here: the work as perceived consists of layers of mediated objects dynamically producing trajectories between each other. This process proposes a re-articulation of the relationship between video, sound, performance and presence. And it suggests a way of engaging with music in general: as matrix of memories across different time-scales, internal and external to the present performance; as the product of an untraceable pattern of nested, folded inscriptions and reencodings.

The full installation requires considerable resources to be mounted. At its heart, however, are the scores for viola d'amore and alto flute – four movements in both cases, each of which passes through the original material (a structural ghost, never heard in its simple form) with a different polyphony of layers of memory. These items (*Apollo 1-4, Marsia 1-4*) have been prepared for separate performance in the form of a score + Max patch.

Acknowledgements. A vital lesson of this project has been the importance of teamwork, of the role of distributed creativity in the production of such technologically hybrid work. Heartfelt thanks are due to Richard Craig (alto flute) and Marco Fusi (viola d'amore), Shivadas De Schrijver (videographer), Juan Parra Cancino (sound engineer), Magno Caliman and Elisabeth Damai Salverda (instrument builders) and Leonardo Impett (coding).

References

Budson, Andrew E., Kenneth A. Richman, and Elizabeth A. Kensinger. 2022. "Consciousness as a Memory System". *Cognitive and Behavioral Neurology,* 35: 263–97.

Deleuze, Gilles. 2013. *Cinema II: the Time-Image.* Translated by Hugh Tomlinson and Robert Galeta. Bloomsbury Academic.

Huron, David. 2008. *Sweet Anticipation: Music and the Psychology of Expectation.* MIT Press.

Kahana, Michael J. and Jonathan F. Miller. 2013. "Memory, Recall Dynamics". In Harold Pashler (editor) *Encyclopedia of the Mind.* Volume 2. SAGE Publications. 492–496.

Mainstone, John. 1998. "The physics of the organ". In Nicholas Thistlethwaite and Geoffrey Webber (editors) *The Cambridge Companion to the Organ.* Cambridge University Press. 37–41.

Noë, Alva. 2015. *Strange Tools: Art and Human Nature.* Hill and Wang.

Ovid. 1986. "Book VI" in A.D. Melville (translator) *Metamorphoses.* Oxford University Press. 121–143.

Pöhlman, Egert, and Martin L. West. 2001, *Documents of Ancient Greek Music.* Clarendon Press.

Reinach, Théodore. 1893. "La musique des hymnes de Delphes". *Bulletin de Correspondance Hellénique.* Vol. 17. 584–610.

Snyder, Bob. 2016. "Memory for Music". In Susan Hallam, Ian Cross, and Michael Thaut (editors) *The Oxford Handbook of Music Psychology*. 2nd edition. Oxford University Press. 167–180.

Stiegler, Bernard. 2010. T*echnics and Time, 3: Cinematic Time and the Question of Malaise.* Translated by Stephen Barker. Stanford University Press.

Tomé, Douglas F., Ying Zhang, Tomomi Aida et al. 2024. "Dynamic and selective engrams emerge with memory consolidation." *Nature Neuroscience* 27, 561–572. https://doi.org/10.1038/s41593-023-01551-w

Viola, Bill, 2013. "The Sound of One Line Scanning". In Dan Lander and Micah Lexier (editors) *Sound by Artists*. Charivari Press. 39-62.