Sonic and Visual Relationships: The Mind, Contemporary Composition and Complex Signals

Luís Arandas

Faculty of Engineering, University of Porto – INESC-TEC

luis.arandas@inesctec.pt

Mick Grierson

University of the Arts London – Creative Computing Institute

m.grierson@arts.ac.uk

Miguel Carvalhais

Faculty of Fine Arts, University of Porto – INESC-TEC

mcarvalhais@fba.up.pt

INTRODUCTION

Considering synchresis[1] and the possible relationships that sound and moving images may have is something deeply rooted in artistic practice. This is a topic that goes back to classical Greek perception theories of seeing and hearing (Darrigol 2010), and can also be found in Newton's work on harmonies of sound and color (Briggs 2007). As posed by Ribas (2011), several artists and thinkers have sought to create both experiences and systems that intersect these two fields. As examples we can mention artworks where video[2] is orchestrated by sound (Kurokawa 2011), or where visuals are sonified (Yeo and Berger 2008), as well as the treatment of both media as a whole. This latter is a demanding journey, one in which Adriano Abbado (1988) succeeded by establishing links between audio and video dynamic events proposing a specific model for image-sound correspondence.

The drive of such willingness can come from different places, because one feels comfortable expressing it that way, or even through mere exploration. Still, sometimes outside the field of arts, there is a great enthusiasm to understand visual and auditory phenomena consistently. To correlate frequencies, to measure perceptual reactions and even to find almost impressionistic meetings between video and music.[3] Science has been measuring these activities using pragmatic procedures and conditions. We try to understand how nature behaves, just as we try to understand it in our minds. We pose questions such as: How and why does this type of experience, coming from this type of artwork and system, happen in me? This can clearly be asked, both scientifically and artistically, and we argue here that this is one of the places where art, philosophy and neuroscience might come together.

COMPOSITION AND MEMORY

In composition we strive to model sound and video, both in the artwork and in the systems used. Compositional processes, as the root of artistic practice, exist in the materialization of intersubjective ideas, even if

[1] Defined by Michel Chion as a portmanteau of *synchronism* and *synthesis*, as the "spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time" (1994). Used here as a term that describes something quite present in artistic practice for mixed media since the twentieth century.

[2] Whenever video is mentioned in this document it is intended to represent moving images (e.g. procedural graphics or film).

[3] *Impressionist* meaning has to do with qualitative opinions one might create, we do not disregard that types of structures in the artwork can be quantified and objectively induce different states of experience. For terminologies such as *color music* and advocates of abstract analogies between music and visual art see (Zilczer 1987).

transformationally. These processes, and promoting the multidisciplinary discussion raised, are linked with episodic memory,[4] consisting of *episodic elements* where records of experience as visual images associated with concepts[5] appear (Conway 2009).

What happens in our minds is quite subtle and knowing how it happens further inspires artificial intelligence and computer science research. We don't remember entire pieces we just heard in the theatre, or the entire surface of paintings we found in the nearby museum. But we do remember symbols and patterns in those same pieces and paintings that last in our memory and are dear to us.[6] It is subtle, but we eventually take it in consideration when making algorithms, even unconsciously. The abstract, discrete systems we build to make art, when modelled like us, can help us understand a little bit more about the world through the art we are trying to make with it.

COMPLEX SIGNALS AND MIND

Taking into account audiovisual art-making and contemporary compositional awareness, we get to the point where we have to determine how these will happen through the computer. Even though data can be discretized in the same formats,[7] the way we represent audio and video signals on a computer is very different. The emerging patterns that exist from the combination of data flows, as well as the way we work with them, directly create clear boundaries. We may apply the same strategies,[8] such as corpus-based search and synthesis (Takahashi 2018; Mital et al. 2013) or database learning but we must also be careful on the way we prepare the algorithm. Deep networks are very different in this matter, even though the symbolic principle can be practically the same.

Where this gets interesting, is when we find emerging phenomena that we can recognise as identical or synergetic. We then raise two topics of discussion:

- The study and modeling of how signal complexity correlates across spectral audio and spectral video frames.
- The study of how the correlations raised can guide the development of new systems and connect with human perception and memory.

Sometimes the detection of similarities between audio and video components can deviate from these objectives. For example, the nature of the similarities, could not be useful for us. Due to the complexity of this subject and the numerous applications of multimodal analysis techniques, there are several reference axioms. One interesting point of view, more efficient than just correlating audio energy and pixel values is posed by Monaci et al. (2007), with the complexity of the multimodal fusion algorithms concentrated on the modelling of modalities, so that *meaningful* structures can be extracted from the signals and synchronous patterns detected (see also Olivier et al. 2007). Regarding composition, the qualitative analysis of these experiments can offer an interesting space for reflection, which in turn fosters mathematical improvement.

By focusing on the results we have, as well as on the techniques (Zhu et al. 2020), we are able to build system architectures specialized in producing these similarities. As for the human brain, as pointed out by Fujisaki et al. (2004), there is an attempt to adjust subjective simultaneity across different modalities by detecting and reducing time lags between inputs that likely arise from the same physical events. Perceptual factors like these, when taken into consideration, may guide us in choosing the right techniques we employ for the development of systems, which in turn will be mirrored in the artwork that comes out of them.

REFERENCES

Abbado, A. (1988) "Perceptual Correspondences of Abstract Animation and Synthetic Sound." PhD diss., Massachusetts Institute of Technology (1988).

[4] The word *memory* being a way of objectifying an entity almost *external* to us capable of producing symbols that eventually appear in our minds.

[5] This article also mentions that episodic memories are embedded in a more complex conceptual system in which they can become the basis of autobiographical memories (Conway 2009). Also, when mentioning mental *images* here, are of type visual. For auditory imagery see Hubbard (2010).

[6] E.g. musical resolutions, bridges, tonal spaces, color, friction, shape.

[7] Relating to the argument that says that everything in a computer is exactly the same: *bit streams*.

[8] In complex digital systems the same strategies can also have different ways of execution, also accompanied by different process hierarchies.

Briggs, D. (2007) "The Dimensions of Color. Modern Color Theory for Traditional And Digital Painting Media". huevaluechroma (blog), viewed 6 August 2020, <https://huevaluechroma.com/index.php>.

Chion, M. (1994) *Audio-Vision: Sound on Screen*. New York: Columbia University Press.

Conway, M. (2009) 'Episodic memories', *Neuropsychologia* 47.11 (2009): 2305-2313.

Darrigol, O. (2010) 'The analogy between light and sound in the history of optics from the Ancient Greeks to Isaac Newton. Part 1.' *Centaurus*, 52.2, 117-155.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. 'Recalibration of audiovisual simultaneity' *Nature neuroscience,* 7.7, (2004): 773-778.

Gillet, O., Essid, S., & Richard, G. (2007) 'On the correlation of automatic audio and visual segmentations of music videos' *IEEE Transactions on Circuits and Systems for Video Technology* 17.3 (2007): 347-355.

Hubbard, T. (2010) 'Auditory imagery: empirical findings' *Psychological bulletin,* 136(2), 302.

Mital, P., Grierson, M., & Smith, T. (2013) 'Corpus-based visual synthesis: an approach for artistic stylization' *Proceedings of the ACM Symposium on Applied Perception*.

Monaci, G., Jost, P., Vandergheynst, P., Mailhe, B., Lesage, S., & Gribonval, R. (2007) 'Learning multimodal dictionaries' *IEEE Transactions on Image Processing,* 16.9 (2007): 2272-2283.

Ribas, L. (2011) 'The Nature of Sound-Image Relations in Digital Interactive Systems' PhD thesis. Universidade do Porto Faculdade de Belas Artes, Porto.

Takahashi, K. (2018) 'The development of corpus-based computer assisted composition program and its application for instrumental music composition.'.

Yeo, W. S., & Berger, J. (2008) 'Raster scanning: a new approach to image sonification, sound visualization, sound analysis and synthesis' *Proceedings of ICMC*.

Zhu, H., Luo, M., Wang, R., Zheng, A., & He., R. (2020). 'Deep Audio-Visual Learning: A Survey' *arXiv preprint arXiv:2001.04758*.

Zilczer, J. (1987) 'Color Music: Synaesthesia and Nineteenth-Century Sources for Abstract Art' *Artibus et historiae*, 101-126.

ARTWORKS

Kurokawa, R. (2011) "syn_ - Audiovisual Concert 2.1ch 30'00-45'00". viewed 6 August 2020, <https://ryoichikurokawa.com/project/syn_.html>-.