

terre: An Online VR Installation for Critical Reflection on Machine Learning and Audiovisual Media

Hugo Scurto

Reflective Interaction group, EnsadLab / EnsAD—PSL Research University

hugo.scurto@ensad.fr

ABSTRACT

In this paper, I present *terre*¹, an online virtual reality (VR) installation offering a critical reflection on Machine Learning (ML) and audiovisual media. As a researcher in Human-Computer Interaction (HCI) and ML and practitioner in music and design, I consider *terre* not only as an artistic experimental installation, but also as a scientific experimental setup for interdisciplinary reflection on interaction design for sound and image. Rather than precise analyses, methods, or answers, my wish is to create an interactive installation that would be experienceable by HCI and ML researchers, in the hope of making certain design issues toward audiovisual media sensible and debatable between communities.

INSTALLATION DESCRIPTION

In *terre*, visitors are virtually put inside a cube, whose six faces successively display experimental music videos. Visitors can freely move inside the cube while watching and listening to these music videos using a VR display—or using keyboard and mouse with a personal computer and screen. While the installation is intended to be experienced linearly—watching and listening to each music video sequentially as they automatically start—, visitors can jump non-linearly between music videos using a pop-up menu displayed at the bottom of the web page. Lastly, visitors can experience a “B-side” version of *terre*—listening to six supplementary music tracks while moving outside the cube—using the same pop-up menu. *terre* is implemented with *A-Frame*; as such, a live demo of its current version is experienceable online².

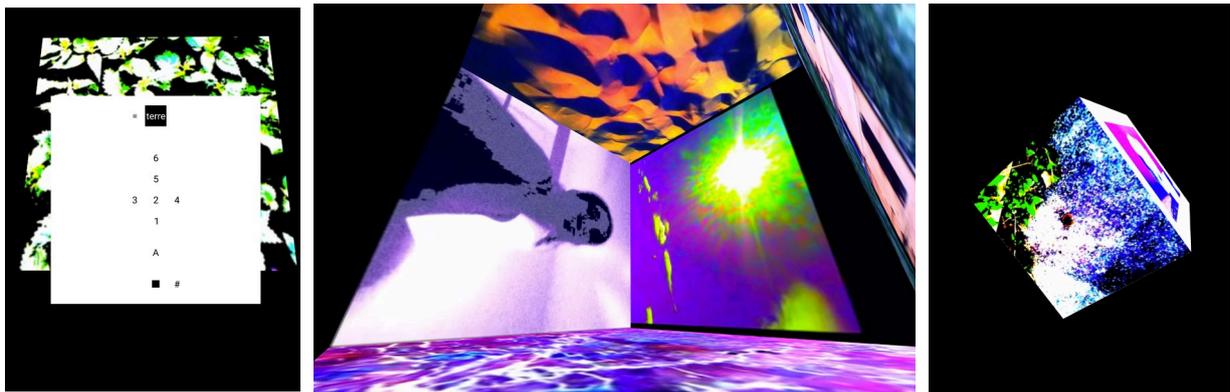


Figure 1: Image excerpts of *terre*: pop-up menu, cube (inside view), cube (outside view).

¹ *terre* exists as the first part of an online ML installation, *ciel = terre*, whose second part, *ciel*, was presented at NIME2020 (Scurto 2020).

² <http://wolfgang.wang/terre>

By having VR visitors simply stare at movie screens and listen to recorded music, *terre* brings the notion of interactivity back to the experience of traditional audiovisual media. As such, it intends to raise three critical points related to HCI and ML research and design with sound and image: (1) the material affordances of audiovisual media, (2) the culturally-dependent perception of digital environments, and (3) the structural intelligence of interactive storytelling.

The material affordances of audiovisual media

In HCI, as well as in ML applied to sound, interactivity is often considered as the main feature to be designed in an innovative system. Yet, it does not always account for how audiovisual media themselves could be perceived as interactive by users. For example, recent works on movement sonification for rehabilitation found that sound and music design was deemed as relevant as ML-based movement-sound mappings to motivate users to perform a task (Bevilacqua et al. 2018).

I was interested in using traditional audiovisual media within the VR environment to make these material affordances more palpable for visitors. Specifically, I used flat screens instead of three-dimensional virtual objects to underline how moving images may already be perceived as interactive by users, following a cinema approach (Manovich 2002). Also, I used binaural recordings as basis for music composition to account for the spatial immersion abilities of natural soundscapes (Truax 1984), to which I added rhythmic synthetic sounds to account for the expressive affordances of music itself (Leman 2016). Lastly, I displayed these media into an elementary cube to reduce the VR space's complexity and emphasize material affordances of sound and image.

The culturally-dependent perception of digital environments

In HCI, especially in the course of current deep learning applications, many creativity support tools based on sound and image started being conceptualized, developed, and experimented, typically providing users with automatic media generation abilities—e.g., music (Briot et al. 2017), or image (Davis et al. 2016). Yet, these tools are often built with minor or tacit cultural analyses of digital data used to engineer them—see the musicological and ethical issues raised by the one-million song and lyric dataset recently scraped by *OpenAI* (2020). As a result, images and sounds generated by such tools tend to be standardized, or else, characterized by some form of new aesthetics typical of Artificial Intelligence (AI) (Pachet et al. 2020).

I was interested in exploring contemporary digital aesthetics to inquire this fast-paced quest for innovative AI applications. Specifically, I intentionally altered images and soundscapes using raw editings to evoke graphics and soundtracks typical of video games from the fifth generation era, following a post-digital approach (Cramer 2015). I also decided to reuse personal visual archives captured with my smartphone for video making to account for how mobile web technology is reconfiguring our contemporary society and its images (Manovich 2016). The fact that different categories of users—e.g., non-video game players, or non-social network users—would perceive these digital environments with different familiarity suggests that cultural and aesthetic factors should be taken into account when designing AI-based HCI tools.

The structural intelligence of interactive storytelling

Many HCI systems developed within this wave of ML applications have focused on enabling users to interact more directly with audiovisual media. For example, interactive machine learning techniques for mapping allowed users to use their movement to interact with sound (Françoise et al. 2014). Deep learning models enabled users to easily navigate through large amounts of pictures by computing smooth interpolations (*RunwayML*, 2020). Despite their strong potential for innovation, most of these models still lack interactive workflows enabling richer user interaction with sound and image over time.

I was interested in using interactive storytelling as a high-level form of structural adaptation, or intelligence (Mateas et al. 2003), which could inspire the design of the so-called intelligence of ML models for HCI with sound (Scurto et al. 2019) and image (Koch et al. 2019). Specifically, I used a pop-up menu—freely inspired by a CD's tracklist and a video game pause menu—to let visitors create their unique story based on the six music videos. Also, I created a B-side version of *terre* not only to make another nod to recorded music formats, but also to push visitors to experience flaws of the interactive installation in their personal way (Cardoso et al. 2013). While familiar to HCI researchers and video games designers—as well as to music producers to some extent—, such narrative aspects could be re-injected in ML models applied so as to empower users in their creative process with sound and image (Parke-Wolfe et al. 2019).

ENVISIONED PERSPECTIVES

At time of writing, *terre* was not shown in public yet. Current perspectives include having the broad public, including researchers and practitioners in HCI and ML, experiment the installation so as to collectively discuss its raised issues. As *terre* will also be featured in SIIDS Art track, the symposium would constitute a first opportunity to examine both the installation and its practice-based approach. In the end, future work will include reflexive analysis of how such a practice-based approach to interdisciplinary research in HCI and ML may complement qualitative and quantitative methods usually employed in interaction design for sound and image.

REFERENCES

- Bevilacqua, F., Segalen, M., Marchand-Pauvert, V., Peyre, I., Pradat-Diehl, P., & Roby-Brami, A. (2018). Exploring different movement sonification strategies for rehabilitation in clinical settings. In *Proceedings of the 5th International Conference on Movement and Computing* (pp. 1-6).
- Briot, J. P., Hadjeres, G., & Pachet, F. D. (2017). Deep learning techniques for music generation--a survey. *arXiv preprint arXiv:1709.01620*.
- Cardoso, P., & Carvalhais, M. (2013). Breaking the game: The traversal of the emergent narrative in video games. *Journal of Science and Technology of the Arts*, 5(1), 25-31.
- Cramer, F. (2015). What is 'Post-digital'?. In *Postdigital aesthetics* (pp. 12-26). Palgrave Macmillan, London.
- Davis, N., Hsiao, C. P., Yashraj Singh, K., Li, L., & Magerko, B. (2016). Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (pp. 196-207).
- Françoise, J., Schnell, N., Borghesi, R., & Bevilacqua, F. (2014). Probabilistic Models for Designing Motion and Sound Relationships. In *Proceedings of the 2014 International Conference on New Interfaces for Musical Expression* (pp. 287-292).
- Koch, J., Lucero, A., Hegemann, L., & Oulasvirta, A. (2019). May AI? Design ideation with cooperative contextual bandits. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- Leman, M. (2016). *The expressive moment: How interaction (with music) shapes human empowerment*. MIT press.
- Manovich, L. (2002). *The language of new media*. MIT press.
- Manovich, L. (2016). Instagram and contemporary image. *Manovich net*, New York.
- Mateas, M., & Sengers, P. (Eds.). (2003). *Narrative intelligence*. J. Benjamins Pub..
- OpenAI (2020), *Jukebox*, OpenAI, viewed 3 August 2020, <<https://openai.com/blog/jukebox/>>.
- Pachet, F., Roy, P., & Carré, B. (2020). Assisted music creation with Flow Machines: towards new categories of new. *arXiv preprint arXiv:2006.09232*.
- Parke-Wolfe, S. T., Scurto, H., & Fiebrink, R. (2019). Sound Control: Supporting custom musical interface design for children with disabilities. In *Proceedings of the 2019 International Conference on New Interfaces for Musical Expression*.
- RunwayML (2020), *Machine Learning for Creators*, RunwayML, viewed 3 August 2020, <<https://runwayml.com/>>.
- Scurto, H., Van Kerrebroeck, B., Caramiaux, B., & Bevilacqua, F. (2019). Designing deep reinforcement learning for human parameter exploration. *arXiv preprint arXiv:1907.00824*.
- Scurto, H. (2020), *ciel*, NIME2020, viewed 3 August 2020, <<https://nime2020.bcu.ac.uk/ciel/>>.
- Truax, B. (1984). *Acoustic communication*. Greenwood Publishing Group.