Possibilising Performance through Interactive Telematic Technology: Mental Dance

Carol Brown, Monica Lim
University of Melbourne
Melbourne, Australia
carol.brown@unimelb.edu.au, monicaml@student.unimelb.edu.au

Abstract

Mental Dance engages audiences in modes of performance that call attention to multiplicitous dimensions of the present. As an interactive improvised dance-sound-tech event, we invite perceiving attention to the mutable present as always already shaped by perception, prior experience, cognition and corporeal change. Neuroscientific research underscores our approach to a digital interface that proposes new relations between audience and performer as a result of repeated COVID-19 lockdowns. Using MediaPipe pose estimation technology to track dancers' movements from webcam feeds, we directed telematic rehearsals and performance of the work on Zoom where dancers in their home environments sculpt and respond to sound in real-time. Constraints such as forced isolation, lack of access to technology and space to move, were embraced to create a new type of collaborative performance where the screen becomes the stage and the interface between movement and sound. This workflow can be used to enable interactive telematic performance where collaborators are unable to be in the same physical space with no specialist hardware requirements.

Keywords

Telematic performance, choreo-sonic, interactive dance, neuroscience, machine learning, embodied sound .

Introduction

Throughout 2020 and 2021, Melbourne, Australia was under strict lockdown for more than 245 days as a result of COVID-19. We could not leave our homes except for very specific reasons. All creative studios, performance spaces, universities and most workplaces were closed. At the time, we were developing an art-science collaboration where neuroscientific research into human cognition was used as conceptual reference for an interactive dance performance using wearable sensors. We were interested in the choreo-sonic and cognitive aspects of the relationship between movement and sound where the movement creates and/or changes the sound, and how this resonated with neuroscientific concepts such as predictive coding and the Bayesian brain. [1]

Lockdowns meant that we could no longer collaborate in the same physical space, nor use wearable sensors such as accelerometers, gyroscopes, biophysical sensors or other devices such as infrared cameras, as the dancers did not have access to the required computer programmes or specialist hardware. To enable ongoing development, we had to harness ubiquitous technology that everyone could access. All collaborators were familiar with video conferencing apps such as Zoom, so using the video feed from Zoom for movement-tracking was identified as the easiest way to build the interactive system. The criterion was not accuracy, but accessibility and ease of use.

Thematic Concept

The project[1] begun in 2019 as an open-ended inquiry into how neuroscientific concepts can inform and be integrated into a creative process. The aim of the research was to delve into the possibilities of diverse combinatory elements – data, dance and sound – to create new forms of expression that generate unforeseen affects supporting and legitimising diverse experiences of mindbody relations. Through various interviews and visits to the neuropsychiatry lab, thematic threads were teased out for further creative exploration. These included current research into neurodiverse cognition for autism spectrum disorder and schizophrenia, and the research technologies used in the lab including EEG, MEG and fMRI. [2] Key phrases from the literature were extracted to inform choreographic scores such as 'changing states', 'oddball' and 'past is present'.

In contrast to the diagnostic language surrounding neurodiversity, we looked at the human impact of living with mental illness, such as that experienced by dancers Vaslav Nijinsky and Lucia Joyce. [3] [4] Although we

[1] Recording of an online performance can be viewed at https://vimeo.com/624934863.

started in the lab, we wanted to end up *inside* the body with all its mechanosensory neurons and limbic emotions.

Collaboration

As inter-disciplinary artistic researchers, we were interested not only in the creative outcome, but the processes used for collaboration and development.

The collaborations were multi-layered, cross-feeding into each other in an iterative process over 18 months. Ideas formed in one session were taken up in other directions to create a matrix of material-data combinations. Figure 1 illustrates the collaborative links formed between all participants and their areas of mutual influence.

For example, references to Nijinsky's iconic gestures in the choreography became an impetus to use Stravinsky's *Rite of Spring* polychord as a sound sample. Neuroscientific concepts around human learning and predictive coding led to an exploration of unsupervised machine learning to categorise and map sound, resulting in a tandem performance by humans and machines simultaneously learning movement-sound relationships. [5]

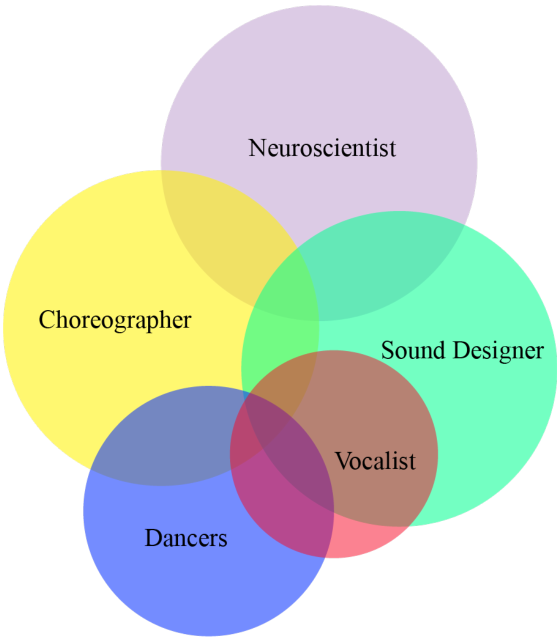

Figure 1. Collaborative links formed between participants.

Performance instructions on breath and phrasing for the vocalist became part of the movement score. Interview transcripts with the neuroscientist became part of the vocal text. Bird sounds used in the EEG oddball auditory test became inspiration for sound samples and movement gestures. The responses of the dancers and the words used to describe their experiences were recycled and reframed into movement and sound ideas. This process of forming and becoming through cross-modal experiences led to an emergent outcome shaped by all participants.

Project Design

Telematic Technology

Rehearsals and performances were held over Zoom. Video feeds of the two dancers were screen-grabbed in TouchDesigner and rescaled to half their original resolution to improve performance.

TouchDesigner was then used to implement a Python script to run MediaPipe on each video feed. MediaPipe is an open-source framework by Google for building cross-platform machine-learning pipelines. [6]

In this case, we experimented with MediaPipe's Face, Hands, Pose and Holistic (which combines Face, Hands and Pose) models. Holistic is the most comprehensive pipeline, enabling tracking of 543 landmarks on the body (33 pose landmarks, 468 face landmarks and 21 hand landmarks per hand). However, running two holistic pipelines in real-time was computationally expensive, leading to an unacceptably low framerate. Tracking fine detail of the face and hands was also less important for this project than overall body landmarks, so the decision was made to use the Pose model only (see Figure 2 and Figure 3).

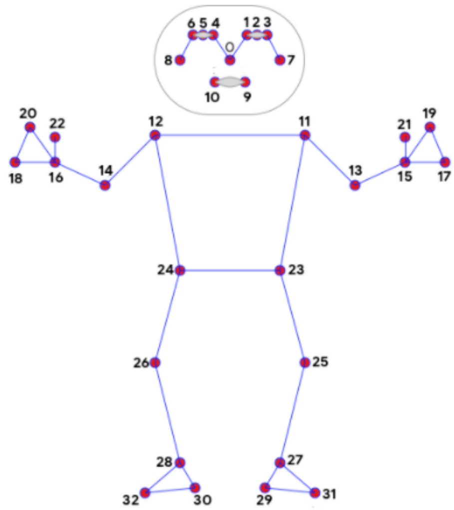

Figure 2. Pose Landmark Model © 2020 Google LLC

Figure 3. MediaPipe Pose landmarks on Zoom video feed.

The Python script was adapted from code released by MediaPipe and the TouchDesigner implementation by Bryan Chung. [7] Selected landmark co-ordinates and joint angles were sent via OSC to MaxMSP for real-time sound processing. Audio output from MaxMSP was streamed to all collaborators and audiences on Zoom using the Loopback virtual audio device.

Choreography

Choreographically, the performance evolved through movement scores that became increasingly precise as the dancers and co-directors continued to develop and refine these during rehearsals. The dancers' sensorimotor skills and corporeal imaginations were continually being tested by the sound and the words being sung in the vocals. Their interactions with the sound relied on perceptual feedback and feedforward loops. Relations to pitch, timbre, voice, text, tempo and tone shaped and structured movement and action within a choreographic dramaturgy that was developed across four states or scenes: *Neural Networks*, *Noisy Voices*, *Lucia* and *In My Head*. Given the unfamiliarity of the dancers with the telematic environment they were working in, and the fact they could not see or sense each other (kinaesthetically, haptically or through breath exchanges), we relied upon coming to know the structure of the improvisation and referencing the gridded image of their partner through the screenic interface. This shaped a sense of parallel worlds that across multiple rehearsals grew to become intuitively relational. The multi-dimensional and multi-sensory habitus of the dancers performing in their home environments contrasted with the screenic interface with its two dimensional rectilinear framing. Choreographically, we played with proximity and distance, the tactile and the remote, opening vectors of relation across physical and virtual dimensions and pushing the apparatus to behave differently beyond the norms of the Zoom platform. Both dancers, in becoming habituated to each other's movements across rehearsals, developed a synergy which enabled them to shape the choreography from inside the system. We related this to the theme of neural networks as if the 'brain' of the dance was being lived in the sphere of these more than human relations.

Sound Design

The sound design had 4 priorities:

1. Using live vocals as a primary sound source.
2. Ability to convey conceptual links to the neuroscientific ideas.
3. Ability to respond to and express movement.
4. Capacity for dramaturgical development.

The emphasis on using live vocals arose from the impetus to research perceptual feedback loops between dancers and vocalist in addition to any digital feedback systems, allowing more layers of interaction and novelty to emerge. This interest in the *human* working in tandem with the *machine* resonated with the conviction that advancing human betterment should be a central concern in all neuroscientific and technological advances, whether that be research into consciousness, neurodiversity, or Artificial Intelligence. The field of interactive sound-design is also dominated by digitally synthesised sound, and we wanted to expand the ways in which the classical and acoustic traditions could engage with new technologies. The use of Baroque countertenor technique in the vocal score (e.g. ornamentation) coalesced with granular synthesis controlled by movement, collapsing binary distinctions between human and non-human, old and new.

In developing text for the vocal score, phrases from Nijinsky's diaries were combined with scientific language from the literature and interview transcripts using a Markov chain generator. This produced unexpected combinations of technical and intimate language, a kind of technologically-mediated poetry. Selected text (e.g. 'deficit following brain', 'efficient white matter') was also used to improvise and develop movement.

However, the use of vocals as a primary sound source introduced constraints when designing the sound's interactive responsiveness to movement. These constraints include limited frequency range and timbre, as well as inconsistent gain and timing. Therefore, subtractive synthesis or spectral processing techniques were not effective. Instead, granular synthesis was used extensively for vocals, and this was combined with other digital processing techniques such as harmonisers and delay lines. Additional sound sources were added to create more layers of interaction with movement as well as to support dramaturgical development. These included sampled sounds manipulated through concatenative synthesis and filtered waveforms. Although the audio output was limited to stereo on Zoom, binaural sound spatialisation was implemented to accentuate movement.

Mapping

Mapping design has been extensively researched and shown to be vitally important to how an interactive system engages its participants, as well as the audience. [8] Whilst the focus of our sound design prioritised the integration of live sound and thematic concept, effective mapping was still required to give the dancers a sense of agency and control, and sufficient transparency for the audience to understand the movement-sound relationships.

Mapping movement using MediaPipe from a single webcam represented particular challenges. Critically, the depth (or z-axis) in the Pose model is an experimental value based on the depth of a particular landmark relative to the user's hips, rather than a true 3D value. [9] Therefore, it was difficult to compensate for the dancer's distance to the camera in analysing movement. The same movement or gestures produced wildly different results depending on the depth and orientation of the dancer to the camera.

A number of strategies were employed to mitigate this. Firstly, MediaPipe has the ability to output real world co-ordinates in metres with the origin as a point in between the hips, rather than relative to the screen as its default. Therefore, where the dancers' distance to the screen varied significantly, the world landmarks were used.

Mappings that produced specific effects were also limited to choreographic sections where the dancers' distance and orientation to the screen were consistent. For example, one section of the work allowed the dancers to adjust the gain depending on the distance between their wrists. Putting their hands together turned off the sound completely. This mapping relied on knowing that the dancers would sit down facing the screen for that entire section.

Using joint angles rather than absolute distances or rate of change measurements of coordinates (e.g. velocity, acceleration) also created more flexibility in dealing with differences in scale. However, the same pose with different orientation to the screen also resulted in very different joint angle data.

Therefore, where the dancers moved back and forth across the screen, a more general and forgiving mapping strategy had to be implemented where sound parameters would change with movement but did not rely on particular gestures or orientation. This included changes in filters, granulation parameters or pitch-shifting. This was particularly important where dancers would sometimes go off the screen completely or only had parts of their body onscreen with no facial information, as this would result in MediaPipe outputting no landmark information.

Using machine learning was also effective to create responsive and often unpredictable relationships between movement and sound, without requiring any explicit mapping of specific movement parameters to sound.

One advantage afforded by MediaPipe's Pose model is its ability to predict landmarks even where parts of the body are partially obscured or out of the screen, as long as there is sufficient information to recognise a pose region-of-interest (typically, a face). Therefore, provided the mapping design was flexible enough to utilise outlier values, the sound could still interact with movement where the dancers were only partially onscreen.

Conclusion

The research culminated in two public performances, one on Zoom and the other on Microsoft Teams. Through performance, the interdisciplinary layers of the work were manifest and made accessible to a globally distributed audience. The performers, though dancing at home, noted how performance nerves were present and how the experience of dancing telematically held a similar state of attention for them to traditional staging. We use the term 'empathy machine' to describe the sense of relatedness the tracking of attention through distributed cognition and corporeal activations demanded. The feedback loops and shifting improvised relations between dancers and data generated a constantly evolving expressive terrain that audiences were invited to perceive and respond to. Audience feedback on the performances was overall positive, with most people reporting positive engagement despite the onscreen delivery. The use of domestic spaces received mixed feedback, with some in the audience enjoying the 'intimacy' and 'immediacy' of 'personal 3D spaces', whilst others found the 'home background distracting and diluting the emotional impact'. Many in the audience were intrigued by the technology and its potential for further exploration, both in performance and pedagogy. They were also interested in 'trying to figure out' the mapping design, and the impact on the sense of collaboration where movement by two different dancers in different spaces affected the same sound.

Although our shift to remote collaboration was forced by external circumstances, the exploration of a new mode of working resulted in us viewing the online rehearsals and performances as not just temporary, inferior replacements for the 'in-real-life' version, but interesting and worthwhile in themselves.

Future Development

The telematic workflow described in this paper can be used to create other forms of interaction. For example, audiences can participate to create sound, textual or visual interactions in a workshop, installation or performance context, unrestricted by geography and availability of specific hardware.

Multiple sound and visual controllers can also be networked, creating telematic ensembles. Whilst the movement data does not have the accuracy of mocap and is subject to latency, the system's ease of use and accessibility for anyone with a computer and internet connection make it a useful alternative for generating digitally-mediated human interaction.

Acknowledgments

We would like to acknowledge the financial support of the Creativity and Wellbeing Research Initiative at the University of Melbourne and Science Gallery Melbourne in developing this work.

References

[1] M.I. Garrido, M. Sahani and R. J. Dolan, "Outlier responses reflect sensitivity to statistical structure in the human brain," PLoS Comput Biol 9, 3 (2013): e1002999, https://doi.org/10.1371/journal.pcbi.1002999.

[2] Claire D. Harris, Elise G. Rowe, Roshini Randeniya and Marta I. Garrido, "Bayesian Model Selection Maps for Group Studies using M/EEG Data," *Frontiers in Neuroscience* 12 (2018): 598, https://doi.org/10.3389/fnins.2018.00598.

[3] W. Nijinsky, *The diary of Vaslav Nijinsky / translated from the Russian by Kyril Fitzlyon*, ed. Joan Acocella (New York: Farrar, Straus and Giroux, 1999).

[4] Carol Loeb Schloss, *Lucia Joyce: to dance in the wake* (New York: Farrar, Straus and Giroux, 2003).

[5] Benjamin D. Smith and Guy E. Garnett, "Unsupervised Play: Machine Learning Toolkit for Max," *NIME* (2012).

[6] “MediaPipe”, accessed 1 October 2021, https://mediapipe.dev.
[7] Bryan Chung, “Touch Designer”, accessed 15 September 2021, https://github.com/chungbwc/TouchDesigner.
[8] G. Emerson and H. Egermann, “Gesture-sound causality from the audience’s perspective: investigating the aesthetic experience of performances with digital musical instruments,” *Psychology of Aesthetics, Creativity and the Arts* 12 (2017): 96.
[9] “Pose Detection ML Kit”, accessed 1 October 2021, https://developers.google.com/ml-kit/vision/pose-detection.