

International Conference on New Interfaces for Musical Expression

Sensor mesh as performance interface

Teresa Pelinski

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Author keywords

subtlety, gestural interaction, digital musical instruments, artificial intelligence

Research question/s/problem

This thesis seeks to capture subtle and nuanced gestural interaction in digital musical instruments while preserving richness and flexibility in their musical output.

Context/theory

The emergence of digital technologies has broadened the affordances of musical instruments, yet some of their traditional properties do not seem to fully manifest in digital musical instruments (DMIs). In particular, this thesis seeks to capture subtle and nuanced gestural interaction while preserving richness and flexibility in the musical output. There exist examples of DMIs that capture gestural interaction with a high degree of subtlety but a limited variety in the sonic output (e.g., using an acoustic signal to control a digital resonator); or instruments that have very rich sonic outputs but limited (or complex to achieve) subtle control (e.g., using high-dimensional gesture-sound mappings). This thesis explores artificial intelligence approaches that preserve both qualities: subtlety in the gestural capture and richness in the sonic output while maintaining a manageable degree of control (e.g., intentionality in the performance).

In DMIs, the gestural interface and the sound generator are separated by a digital mapping layer, whereas in acoustic instruments, the sound generator (an acoustic resonator) is usually part of the gestural interface. Much has been written in the NIME literature about ‘expressive’ mappings or interfaces [1][2][3][4]. [5] argue against this notion of expression as a “quantity in the interface”, which they understand is rooted in the western paradigm of instrumental music (e.g., virtuosity in the performance of acoustic instruments). Moreover, calling an instrument *expressive* by its configuration, hardware or software is problematic if we assume being an instrument is *becoming* one by its intra-action, inter-action, and in general, relationship with the performer and the environment [6][7][8]. Furthermore, [5] critique puts into focus the experimental practices in which the meaning (or lack of meaning) of the performance lies outside the instrument and performer. Perhaps a more adequate characterisation of the affective potential of musical instruments is that of ‘control intimacy’ by [9]. Control intimacy is subtle musical control of an instrument: the instrument “must respond in consistent ways that are well matched to the psychophysiological capabilities of highly practiced performers” [9]. [9] also refers to the performer’s

‘micro-gestural movements’ and how these are translated to the sound. Similarly, [10] refers to ‘micro-diversity’ as a measure of “how much a performer can turn a piece into her own or, consequently, how much two performances of the same piece can differ” [10].

As aforementioned, subtlety is a natural characteristic of many acoustic musical instruments: physical reality is continuous, as opposed to a discrete sample triggering, as we observe in many DMIs. Virtuoso performers of acoustic instruments have fine control of the sound produced through their gestures in their musical instruments, and might expect the same from DMIs. Furthermore, subtlety might be not only a means of control but also a means of exploration of the instrument, as in material-oriented practices [7]. The common element between those two practices is the emergence of tacit or implicit knowledge in the performer. Tacit knowledge describes that knowledge that we know but cannot tell [11], also referred to as *know-how* or procedural knowledge. It is the implicit knowledge that is highly situated and resists articulation and codification [12]. In this regard, [13] argues that tacit knowledge is more likely to emerge in acoustic instruments than in DMIs due to the lack of “natural mapping between gesture and sound” [13] in the latter. Furthermore, in DMIs, he argues, “the physical force becomes virtual force; it can be mapped from force-sensitive input devices to parameters in the sound engine, but that mapping is always arbitrary” [13]. On the contrary, [14] speaks of ‘ergotic interfaces’ where the energy continuum between the performer’s gesture and the instrument’s sonic output is simulated if the signals produced by the system are in scale and shape with the energy fed into the system. That ‘natural’ scaling implies that the mapping is not entirely arbitrary, as [13] argued. Moreover, it also implies that a subtle variation in the gesture will result in a corresponding subtle variation in the sonic output.

In this context, enactivism puts the focus on the “necessary and close link between perception and action” [15]. From this perspective, [15] consider how musicians build their mental model of a complex system’s dynamics (in this case, the musical instrument) using prior knowledge of integrated sensorimotor experiences. [15] implement these ideas using acoustic sensing to capture the gestural interaction. In their instrument *PebbleBox* [15], a microphone is embedded in a foam-padded container full of polished rocks. The performer can, for example, put her hand in the box and stir the stones. The acoustic signal captured by the microphone is passed through a model that extracts two control parameters for a granular synthesis system. The *PebbleBox* is an example of an instrument that captures gesture with its nuances and subtleties; however, since much of the transduction from gesture to sound occurs

in the acoustic domain, there is not much flexibility or variety in the sonic output. Using [10]’s [10] terminology, the PebbleBox has a high micro-diversity (performance nuances) but a low mid- (performance contrasts) and macro-diversity (stylistic flexibility).

Other DMIs are more versatile and richer in their sonic outputs. For instance, digital keyboards have more dimensions of control (one pitch and velocity per key), which enables them to play in different contexts or styles. However, those dimensions are independent, which might not always be desirable. Namely, three independent parameters are captured in the Yamaha WX7 wind MIDI controller: breath pressure, lip pressure and fingering configuration. Nevertheless, experienced woodwind performers complained that the complex behaviour of a wind instrument was not well represented [1] since the airflow through the reed of a single-reed instrument is a function of the pressure across the reed. In other words, the three variables (breath and lip pressure, fingering configuration) were independent in the WX7 controller but *cross-coupled* in acoustic single-reed instruments [16]. Contrarily, coupling too many dimensions might result in a control space too difficult to navigate, where the performer might not be able to perform with intentionality, or intentionality might be too complex to achieve.

The existent approaches for capturing gestural expression in DMIs struggle to find a balance between subtlety in the interaction and richness and flexibility in the sonic output. In this context, this thesis seeks to capture subtle and nuanced interaction without constraining the flexibility and variety of the sonic output, while keeping a ‘navigable’ control space for the performer.

Methods

We propose the sensor mesh as an interface that captures gesture with its nuances and subtlety, yet without constraining the sonic output’s richness. To capture the nuances in the gesture, the sensor mesh will hypersample the gesture with dozens of vibration sensors spread across the digital musical instrument’s interaction surface. Each sensor will ‘look’ at the gesture from a slightly different perspective.

Nevertheless, dealing with a large number of signals is challenging from the point of view of embedded and real-time systems. The sensor mesh will be connected to a selected embedded platform, which currently affords eight channels of 16-bit analogue inputs and outputs sampled at audio rate. More than one embedded board could be used to process the signals from the sensor mesh, effectively dividing the sensor mesh into sub-meshes. However, this approach implies communicating in real-time across

boards to deal with the interaction as a ‘whole’. Moreover, by processing each sensor signal independently, there is no meaning extracted from the gesture. The sensor mesh should be able to interpret these signals as projections of a high-dimensional gestural interaction and effectively act as a bottleneck of interaction.

Rather than processing the sensor’s signals individually, a more interesting approach (and the one I will follow in this thesis) is the usage of deep learning techniques to reduce the dimensionality of the signals. Since, as aforementioned, the sensors record the same interaction from different perspectives, and due to the high sampling rate, the signals will be notably redundant. Nonetheless, the dimensionality reduction is not trivial (e.g., downsampling) since the gesture should be understood as a ‘whole’ rather than as independent perspectives of a single event. In addition, it should be noted that the signals will have a strong time dependency due to their vibrational nature, however, long analysis windows can not be used for their analysis due to the stringent real-time requirements. I presume that a significant effort in this thesis will be reducing the dimensionality of these signals and finding efficient deep learning techniques¹ to interpret these gestures while keeping their subtleties and nuances.

Using artificial intelligence to forge the sensor mesh’s physical-digital integration is not only sensible from the point of view of dimensionality reduction but also to avoid the encoding of musical theory in the system. In systems where the interaction is explicitly mapped, a particular set of gestural input parameters might be associated with a pitch. There is no ambiguity in the sonic output: that set of parameters will always correspond to the same note. However, in acoustic musical instruments, an ambiguous input might result in an ambiguous output, and performers may ‘push the instruments to their limits’. In a system where the coupling between gestural input and sonic output is not explicit, different performers performing the same exact gestures might interpret the sonic output differently. A non-explicit mapping opens the instrument to interpretation.

Carrying out a deep learning model’s inference step in an embedded platform is challenging not only from the point of view of compiling libraries and frameworks from source but also due to the computational limitations of the environment. Most artificial intelligence used in audio and music involves the usage of large datasets and very intensive models. In a resource-constrained real-time environment, these premises change: the model needs to be as small and efficient as possible to be able to predict the next sample in time. Moreover, common deep learning frameworks such as PyTorch [17] or Tensorflow [18] need to be compiled from source for the processor’s

architecture, which is not trivial due to version mismatches. Moreover, PyTorch does not have any optimisations for the selected embedded platform processor.

Expected outcomes

From the Doctoral Consortium, I hope to get some feedback on my conceptual approach to the question of subtlety and flexibility in digital musical instruments. I am halfway through my first year, so my methods are still not well-defined. I would appreciate suggestions especially on how to evaluate the 'degree' of subtlety of the instrument, as well as its output richness. My contributions to the Doctoral Consortium are, at this point, limited to literature suggestions.

Footnotes

1. A possible technique to implement here is deep compressive sensing , a technique that preserves the information rate of the signal at a lower rate than the Nyquist frequency. In this process, signal redundancy is eliminated during the sampling process, resulting in a lower efficient sampling rate . Other possible strategies involve using auto-encoders to find lower-dimensional latent spaces that still preserve the nuances and subtleties of the interaction. [↵](#)

Citations

1. Rovan, J. B., Wanderley, M. M., Dubnov, S., & Depalle, P. (1997). Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance. In *In Proceedings of Kansei - the Technology of Emotion Workshop* (pp. 3-4). [↵](#)
2. Fels, S. (2004). Designing for intimacy: Creating new interfaces for musical expression. *Proceedings of the IEEE*, 92(4), 672-685. <https://doi.org/10.1109/JPROC.2004.825887> [↵](#)
3. Fels, S., Gadd, A., & Mulder, A. (2002). Mapping transparency through metaphor: Towards more expressive musical instruments. *Organised Sound*, 7(2), 109-126. <https://doi.org/10.1017/S1355771802002042> [↵](#)
4. Menzies, D. (2002). Composing instrument control dynamics. *Organised Sound*, 7(3), 255-266. <https://doi.org/10.1017/S1355771802003059> [↵](#)
5. Gurevich, M., & Treviño, J. (2007). Expression and its discontents: Toward an ecology of musical creation. In *Proceedings of the 7th international conference on*

- New interfaces for musical expression* (pp. 106–111). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/1279740.1279759>
6. Hardjowirogo, S.-I. (2017). Instrumentality. On the Construction of Instrumental Identity. In T. Bovermann, A. de Campo, H. Egermann, S.-I. Hardjowirogo, & S. Weinzierl (Eds.), *Musical Instruments in the 21st Century* (pp. 9–24). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-10-2951-6_2
 7. Mudd, T. (2019). Material-Oriented Musical Interactions. In S. Holland, T. Mudd, K. Wilkie-McKenna, A. McPherson, & M. M. Wanderley (Eds.), *New Directions in Music and Human-Computer Interaction* (pp. 123–133). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-92069-6_8
 8. Waters, S. (2021). The entanglements which make instruments musical: Rediscovering sociality. *Journal of New Music Research*, 50(2), 133–146. <https://doi.org/10.1080/09298215.2021.1899247>
 9. Moore, F. R. (1988). Dysfunctions of Midi. *Computer Music Journal*, 12(1), 19–28.
 10. Jordà, S. (2004). Instruments and Players: Some Thoughts on Digital Lutherie. *Journal of New Music Research*, 33(3), 321–341. <https://doi.org/10.1080/0929821042000317886>
 11. Polanyi, M., & Sen, A. (2009). *The tacit dimension*. Chicago ; London: University of Chicago Press.
 12. Kirsh, D. (2009). Knowledge, explicit vs implicit. In *The Oxford Companion to Consciousness*. Oxford University Press. Retrieved from <https://www.oxfordreference.com/view/10.1093/acref/9780198569510.001.0001/acref-9780198569510-e-193>
 13. Magnusson, T. (2009). Of Epistemic Tools: Musical instruments as cognitive extensions. *Organised Sound*, 14(2), 168–176. <https://doi.org/10.1017/S1355771809000272>
 14. Cadoz, C. (2009). Supra-Instrumental Interactions and Gestures. *Journal of New Music Research*, 38(3), 215–230. <https://doi.org/10.1080/09298210903137641>
 15. Essl, G., & O’Modhrain, S. (2006). An enactive approach to the design of new tangible musical instruments. *Organised Sound*, 11(3), 285–296. <https://doi.org/10.1017/S135577180600152X>

16. Hunt, A., Wanderley, M. M., & Paradis, M. (2017). 2002: The Importance of Parameter Mapping in Electronic Instrument Design. In A. R. Jensenius & M. J. Lyons (Eds.), *A NIME Reader: Fifteen Years of New Interfaces for Musical Expression* (pp. 29–44). Cham, Switzerland: Springer. [↵](#)
17. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems* (Vol. 32). Curran Associates, Inc. Retrieved from <https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html> [↵](#)
18. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*. Retrieved from <https://www.tensorflow.org/> [↵](#)