Test

Building and exploring multimodal musical corpora

from data collection to interaction design using machine learning

Federico Ghelli Visi

GEMM))) Gesture Embodiment and Machines in Music

School of Music in Piteå, Luleå University of Technology, Sweden

Mapping Social Interaction through Sound: Conceptual Framework, Technological Development and Real-World Studies

Humboldt-Universität, Berlin, November 27–29, 2020

Overview

- Building a music performance data corpus containing multimodal measures linked to high-level subjective observations
- Co-exploring complexity: Assisted Interactive Machine Learning

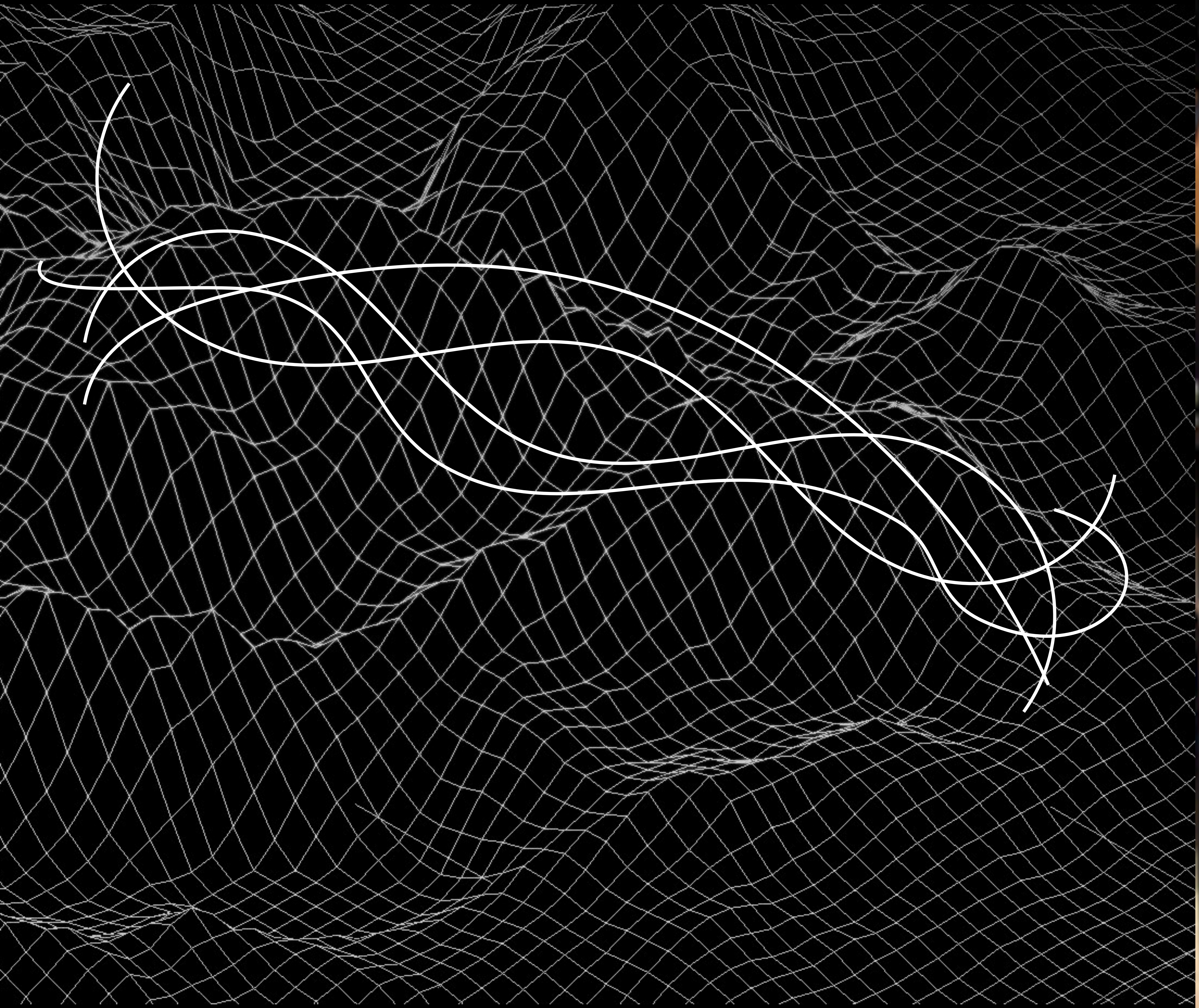
Alban Berg, Vier Stücke Op. 5

Robert Ek: Clarinet

Mårten Landström: Piano

“multimodal”

- Sound
- Movement
- Time and space
- Expressive qualities
- …



A structured multimodal data corpus

Combining qualitative and quantitative data

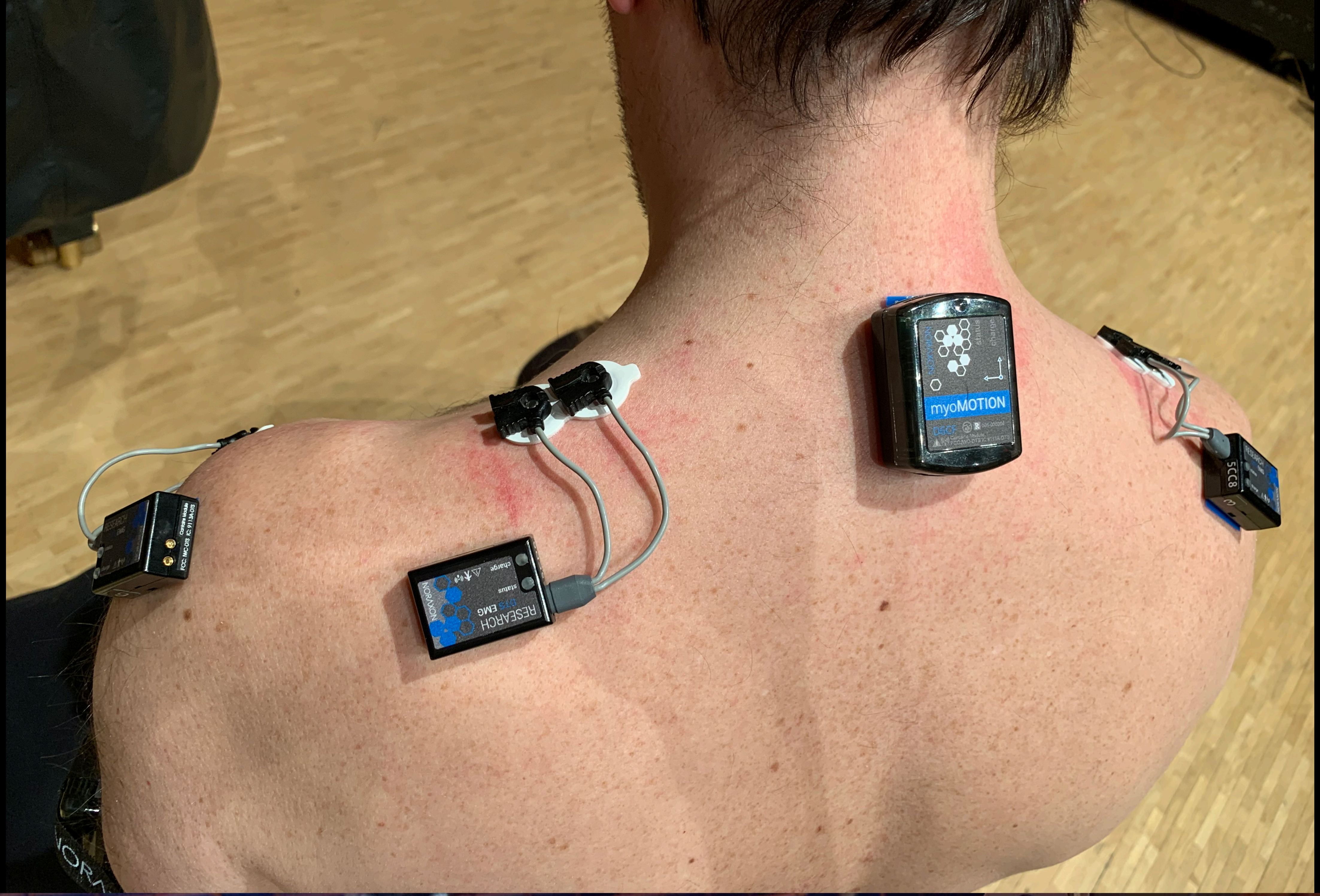
- Not just a “collection” of data but a *corpus of relationships*
- Learn more about the **relationships** between quantitive features (measurements) and qualitative aspect (annotations, expressive qualities…)
- Use labelled and structured data to **understand more about new data** (taking advantage of machine learning approaches for tasks such as automated annotation and segmentation)
- Facilitate the retrieval and analysis of segments and layers of different performances by using low level features as well as higher-level qualities. **Explore similarities, idiosyncrasies**…
- Gain new knowledge on **chunking, coarticulation, and gestural sonic objects**…
- Use the data to enable creative and artistic processes (generative models, multimodal corpus-based synthesis, **live interaction** with sound and other media…)

Data collection
Acusticum concert hall

- Video
  - 2 camera angles
- Audio
  - Clarinet
  - Piano
  - Hall

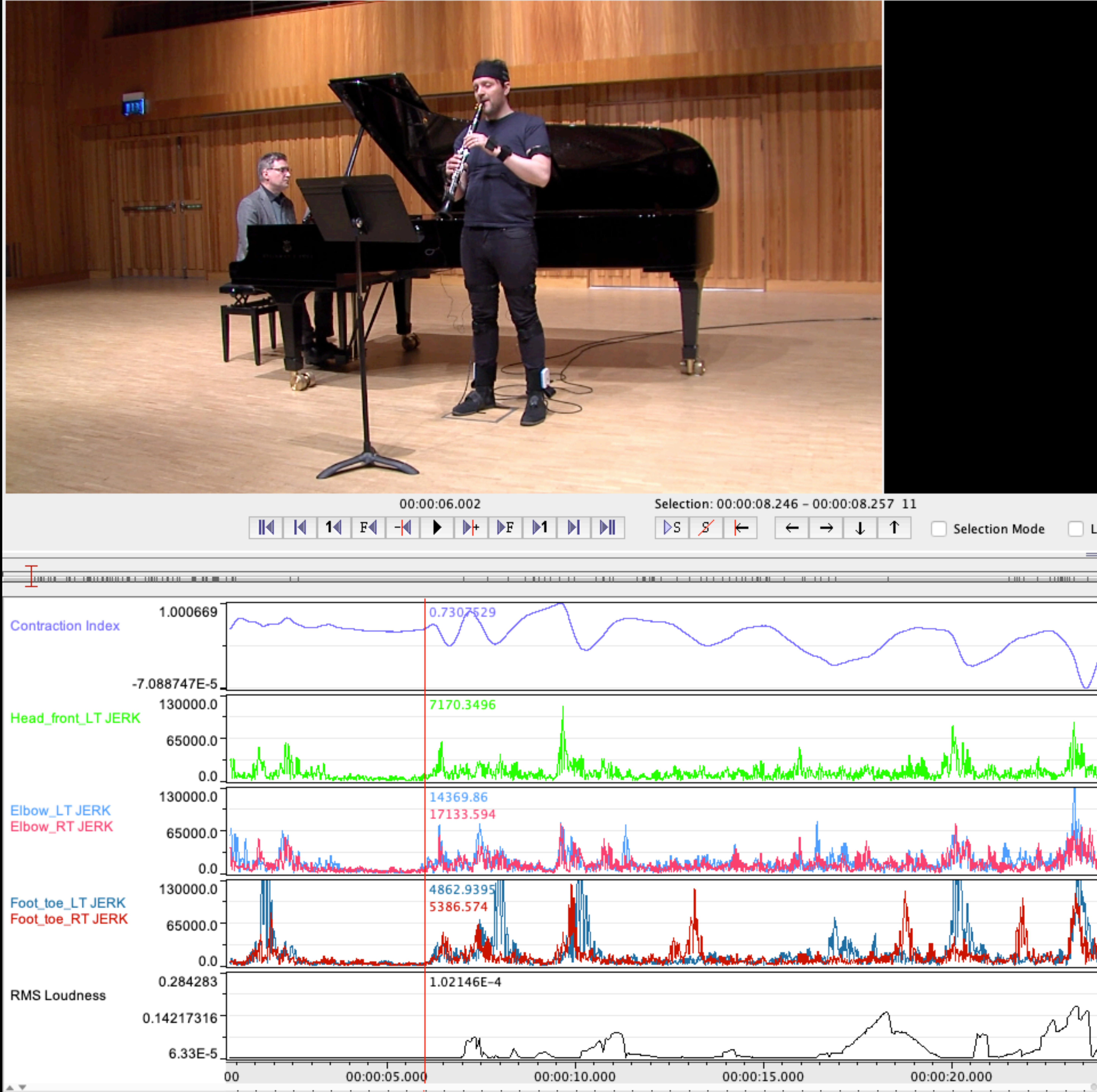
Data collection
Quantitative

- Motion
  - Inertial sensors
- EMG
  - Finger flexors
  - Deltoids
- Insoles
  - Pressure on each foot



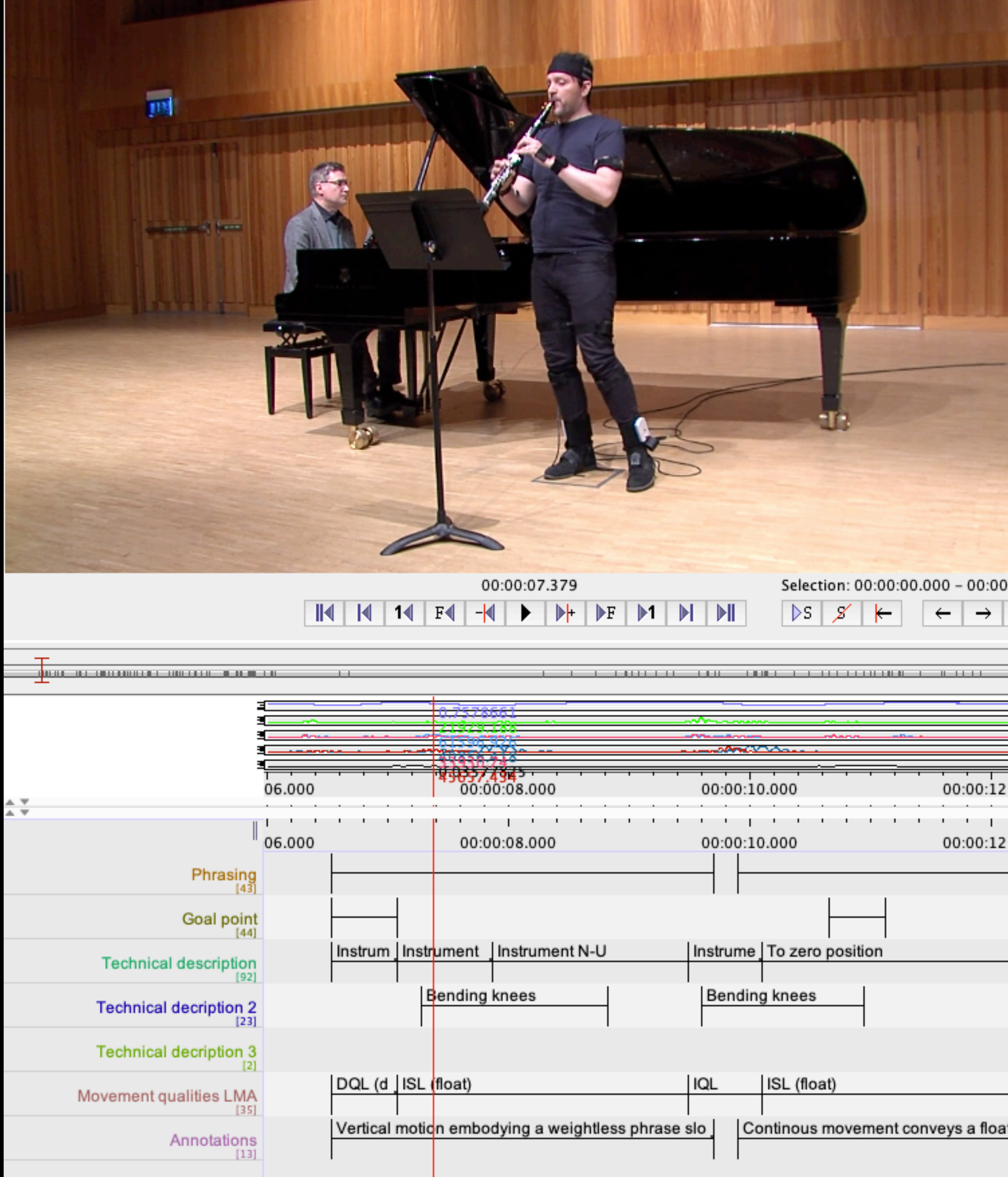
Feature extraction

- Spatial
  - Trajectories
  - Contraction Index
  - …
- Insoles
  - Weight shifting
  - Sum
  - …
- EMG
  - RMS amplitude
  - Onsets and offsets
  - …
- Clarinet audio features
  - Loudness
  - …

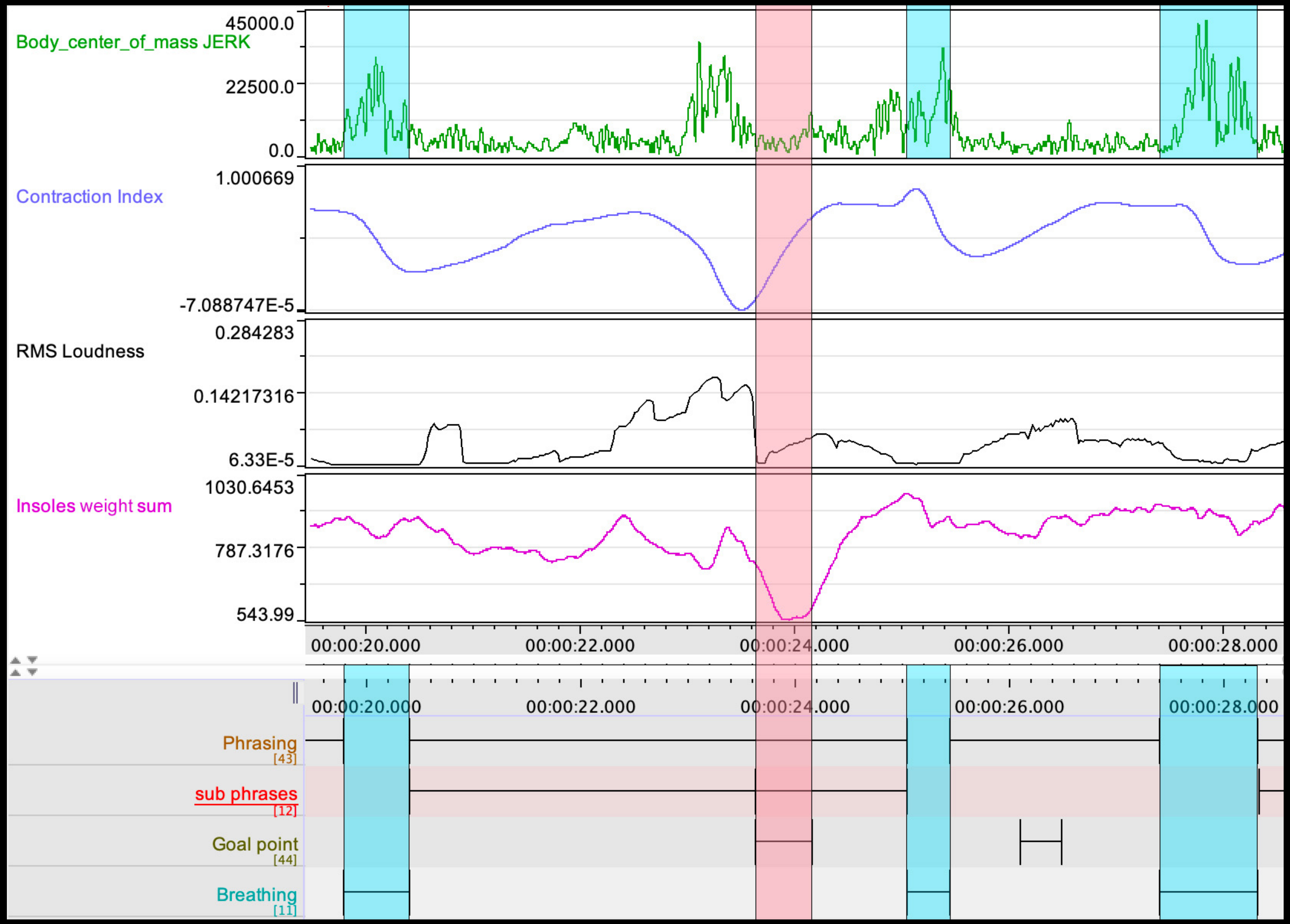


Data collection
Qualitative

- Open coding sessions
- Stimulated recall
- Phrasing
- Goal points
- High-level annotations
- Laban Movement Qualities



Phrasing and Segmentation



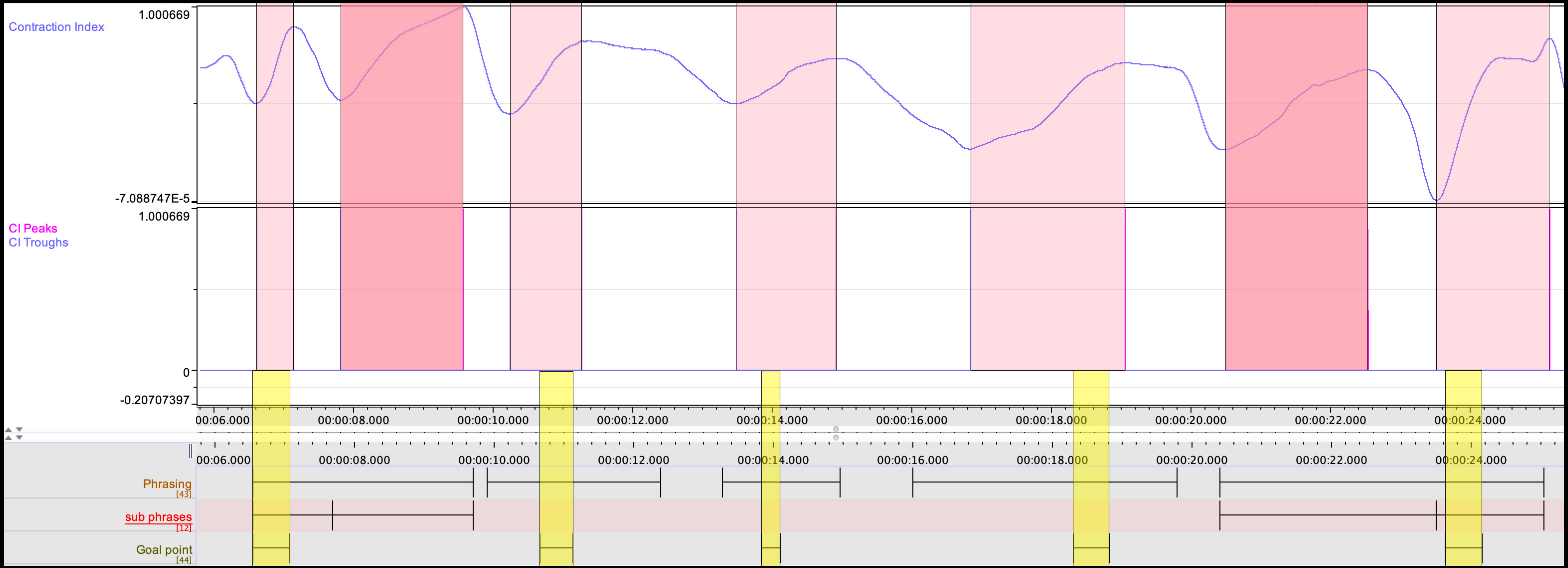
"In the three instances when they coincide with breathing, we see how the peaks in the jerk data coincide with low amplitude in the RMS loudness."

"The second peak in the jerk data, in which the RMS loudness is instead high, does not represent breathing, but rather the performer's preparation aimed at the goal point."

Blue marks annotated breaths and red marks annotated goal point

GEMM

Phrasing and Segmentation



The CI, aligned with the annotated phrasing, and goal points (marked in yellow) in the first 20 s of the first movement. Each rising curve in the CI is marked in red, and the two instances in which the CI does not lead to a goal point are darker.

GEMM

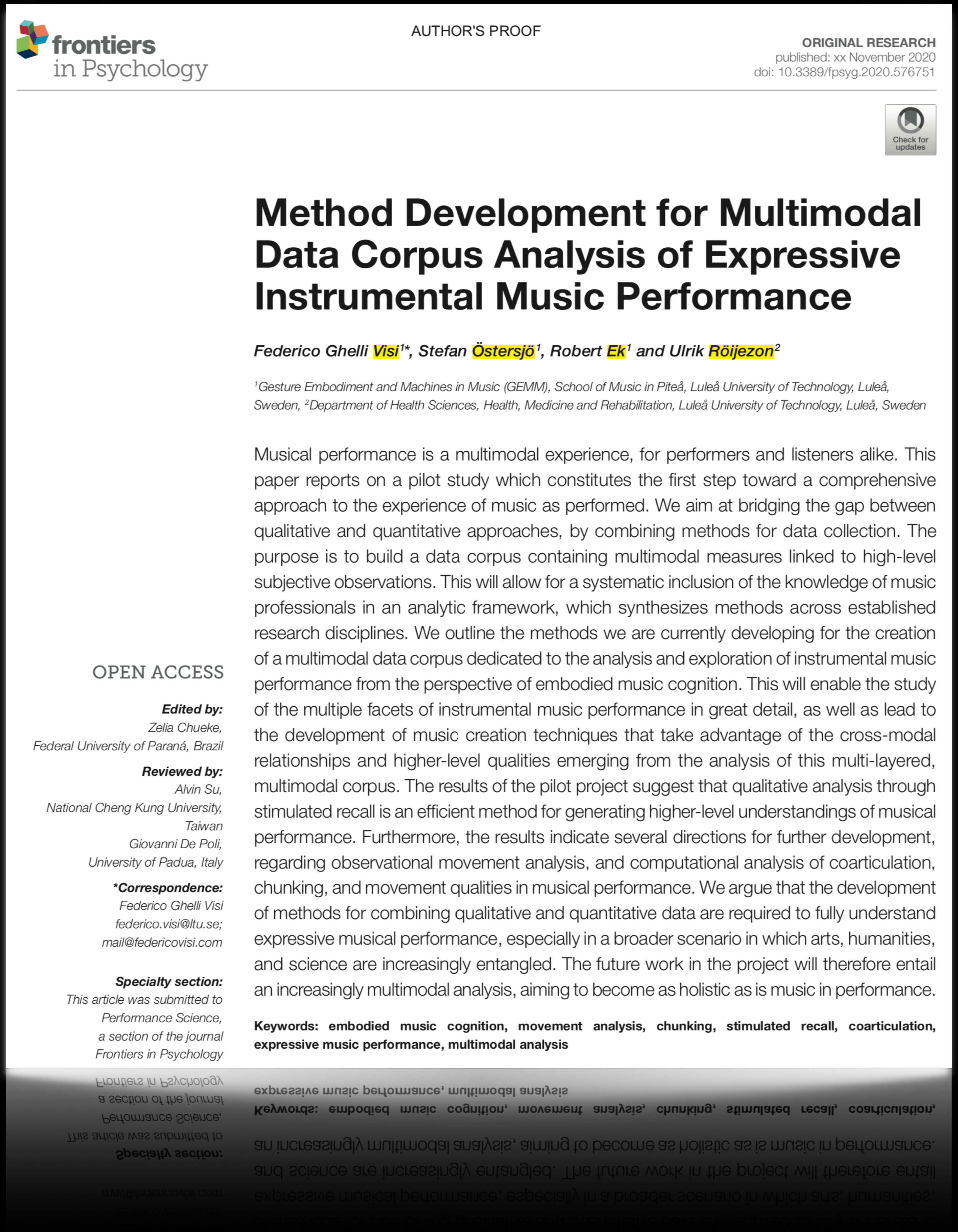
Multimodal concatenative re-synthesis

Contraction index: increasing expansion

References

cutt.ly/visi-mapping-symposium

Visi, F. G., Östersjö, S., Ek, R., & Röijezon, U. (2020). **Method development for multimodal data corpus analysis of expressive instrumental music performance.** *Frontiers in Psychology,* 11(576751), doi: 10.3389/fpsyg.2020.576751



frontiers in Psychology

AUTHOR'S PROOF

ORIGINAL RESEARCH
published: xx November 2020
doi: 10.3389/fpsyg.2020.576751

Method Development for Multimodal Data Corpus Analysis of Expressive Instrumental Music Performance

Federico Ghelli Visi[1], Stefan Östersjö[1], Robert Ek[1] and Ulrik Röijezon[2]*

[1]Gesture Embodiment and Machines in Music (GEMM), School of Music in Piteå, Luleå University of Technology, Luleå, Sweden, [2]Department of Health Sciences, Health, Medicine and Rehabilitation, Luleå University of Technology, Luleå, Sweden

Musical performance is a multimodal experience, for performers and listeners alike. This paper reports on a pilot study which constitutes the first step toward a comprehensive approach to the experience of music as performed. We aim at bridging the gap between qualitative and quantitative approaches, by combining methods for data collection. The purpose is to build a data corpus containing multimodal measures linked to high-level subjective observations. This will allow for a systematic inclusion of the knowledge of music professionals in an analytic framework, which synthesizes methods across established research disciplines. We outline the methods we are currently developing for the creation of a multimodal data corpus dedicated to the analysis and exploration of instrumental music performance from the perspective of embodied music cognition. This will enable the study of the multiple facets of instrumental music performance in great detail, as well as lead to the development of music creation techniques that take advantage of the cross-modal relationships and higher-level qualities emerging from the analysis of this multi-layered, multimodal corpus. The results of the pilot project suggest that qualitative analysis through stimulated recall is an efficient method for generating higher-level understandings of musical performance. Furthermore, the results indicate several directions for further development, regarding observational movement analysis, and computational analysis of coarticulation, chunking, and movement qualities in musical performance. We argue that the development of methods for combining qualitative and quantitative data are required to fully understand expressive musical performance, especially in a broader scenario in which arts, humanities, and science are increasingly entangled. The future work in the project will therefore entail an increasingly multimodal analysis, aiming to become as holistic as is music in performance.

OPEN ACCESS

Edited by:
Zelia Chueke,
Federal University of Paraná, Brazil

Reviewed by:
Alvin Su,
National Cheng Kung University, Taiwan
Giovanni De Poli,
University of Padua, Italy

***Correspondence:**
Federico Ghelli Visi
federico.visi@ltu.se;
mail@federicovisi.com

Specialty section:
This article was submitted to Performance Science, a section of the journal Frontiers in Psychology

Keywords: embodied music cognition, movement analysis, chunking, stimulated recall, coarticulation, expressive music performance, multimodal analysis

Future work

- Assess and consolidate the method for collecting qualitative data
- Apply it to more multimodal recordings (new recording with musician and dancer)
- Share data, methods, and procedures
- Collaboration with AI research group at LTU

  Liwicki, F. S., Liwicki, M., Malo Perisé, P., Visi, F. G., & Östersjö, S. (2020). **Analysing Musical Performance in Videos Using Deep Neural Networks**. In *The 2020 Joint Conference on AI Music Creativity*. Stockholm, Sweden: Royal Institute of Technology (KTH).

- Explore the corpus

Exploring Gesture-Sound Mappings Using Reinforcement Learning

Assisted Interactive Machine Learning

What is it?

<u>In a sentence:</u>

Assisted Interactive Machine Learning (AIML) is an interaction design method based on **Deep Reinforcement Learning** developed for the purpose of **exploring the many possible mappings** between two heterogeneous spaces (e.g. gesture and sound synthesis).

Why?

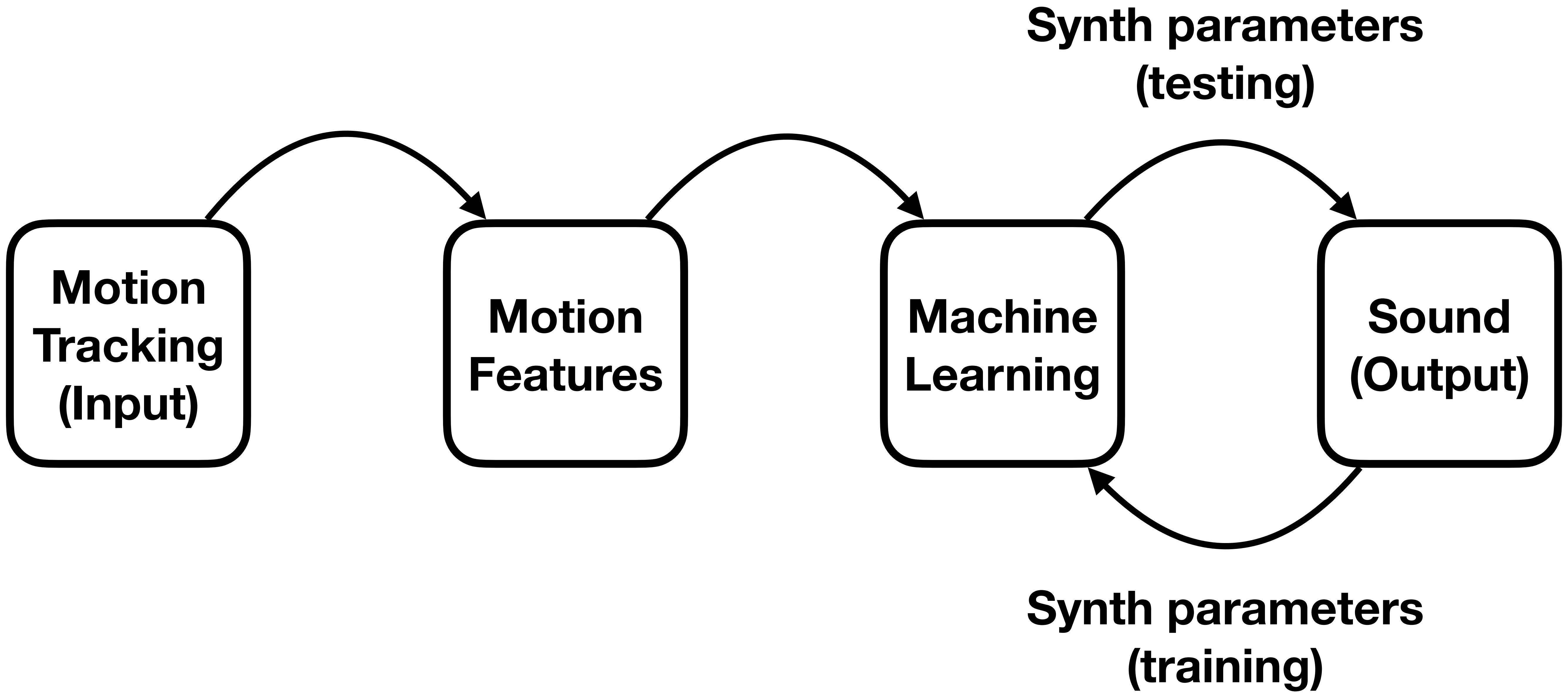
- Machine learning has made designing live interfaces for artistic expression more accessible, allowing rapid creative experimentation and enabling new creative relationships with machines.
- Digital interfaces can be programmed in so many different ways (there are **many possible mappings**).
- **So, what if I could collaborate with an AI to explore mappings I wouldn’t normally think of?**

Why? (cont.)

- Supporting exploration of complex spaces (such as sound synthesis)
- In artistic practice: able to surprise and help breaking creative deadlocks
- Learning: understanding how things work through co-exploring
- Demystifying AI/ML by using it <u>actively</u> to build something

Interactive Machine Learning Model

What: control gestures

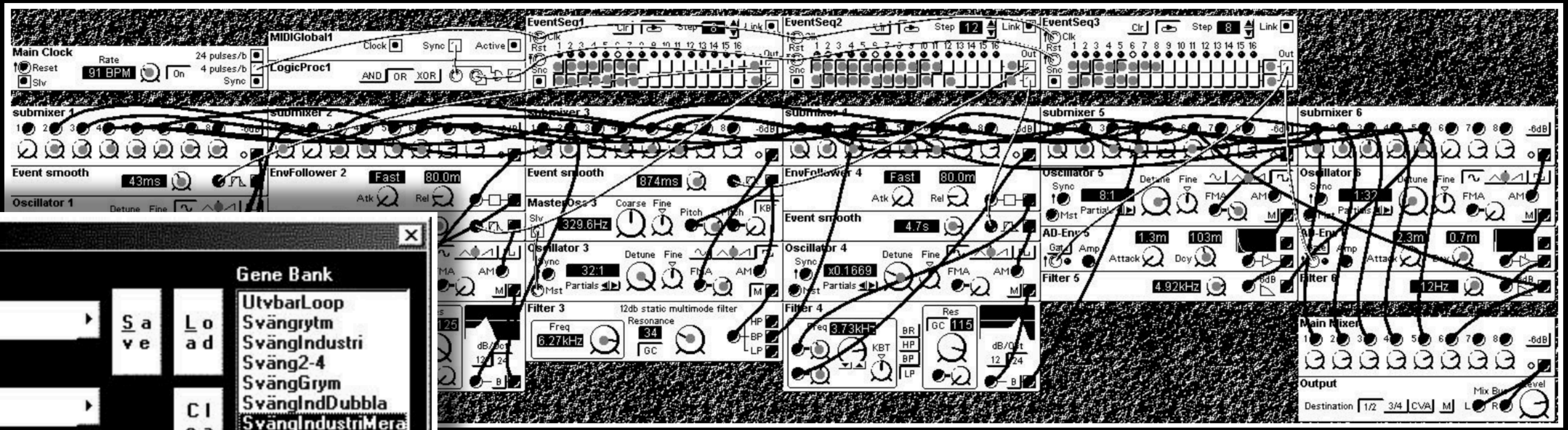


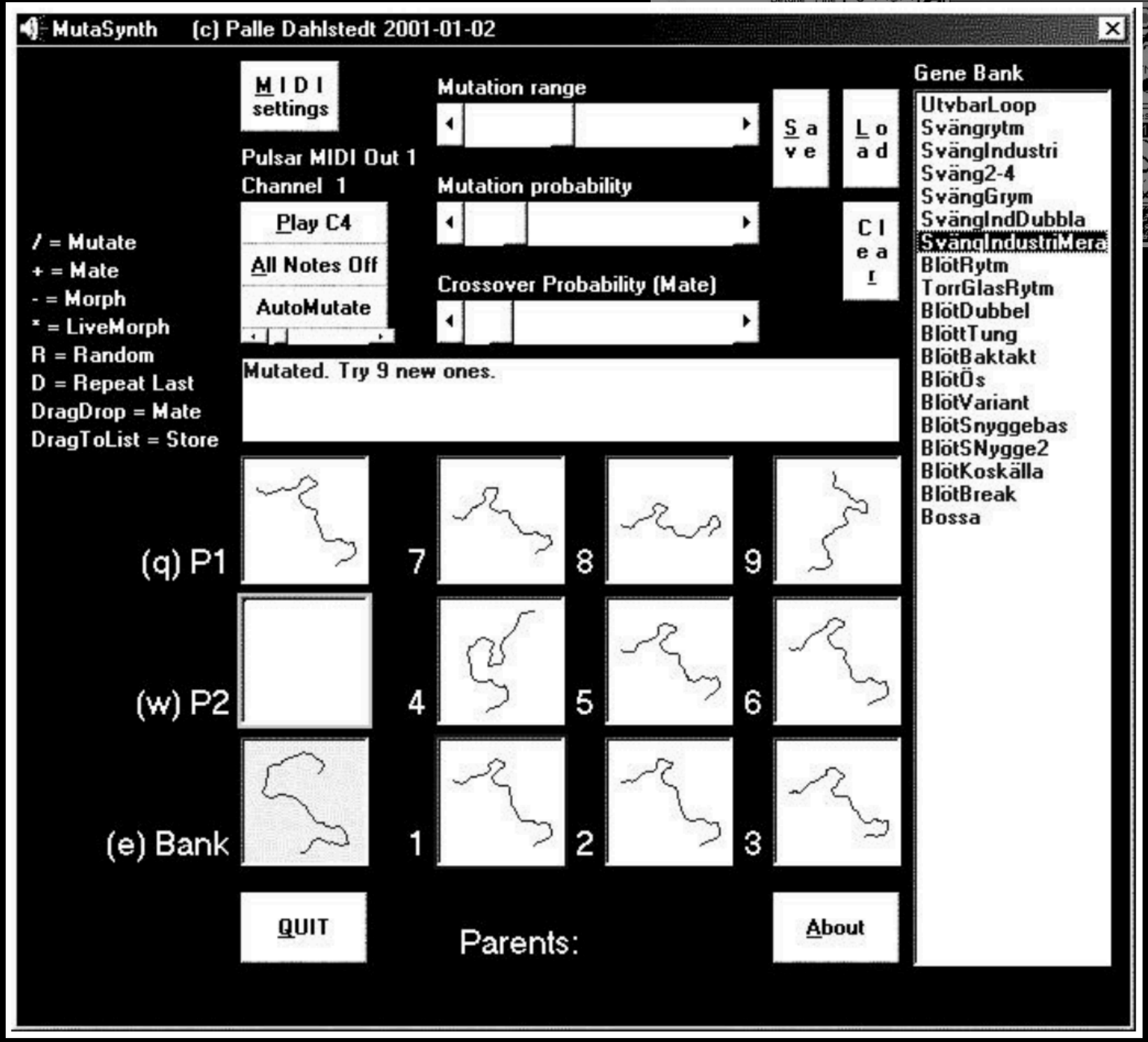
What: sounds

Show the computer how gesture and sound are related by using machine learning

Fiebrink, R. A., & Caramiaux, B. (2018). The Machine Learning Algorithm as Creative Musical Tool. (R. T. Dean & A. McLean, Eds.), The Oxford Handbook of Algorithmic Music (Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190226992.013.23

Exploring Complexity through Emergent Behaviours:
Artificial Agents in Computer Music

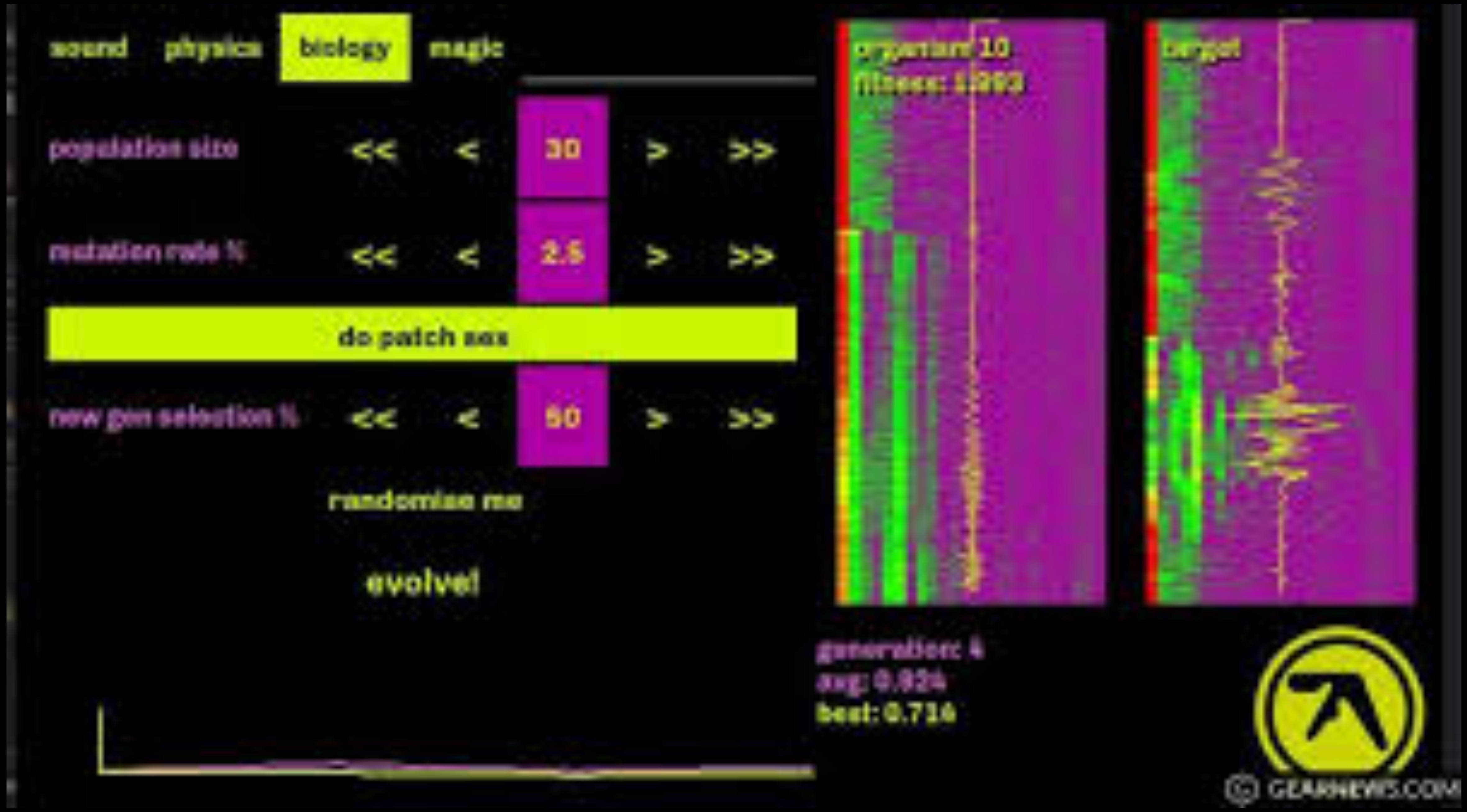




"Interactive evolution as a compositional tool makes it possible to create <u>surprisingly complex sounds</u> and structures in a very quick and simple way, while keeping a feeling of control."

Dahlstedt, P. (2001). Creating and exploring huge parameter spaces: Interactive evolution as a tool for sound generation. In Proceedings of the 2001 International Computer Music Conference.

Genetic algorithms for exploring complex sound parameter spaces



https://fo.am/activities/midimutant/

Review and more recent approaches (also deep networks!):
Yee-King, M. J., Fedden, L., & D'Inverno, M. (2018). Automatic Programming of VST Sound Synthesizers Using Deep Networks and Other Techniques. IEEE Transactions on Emerging Topics in Computational Intelligence, 2(2), 150–159.

Explore Together: Reinforcement Learning

Reinforcement Learning

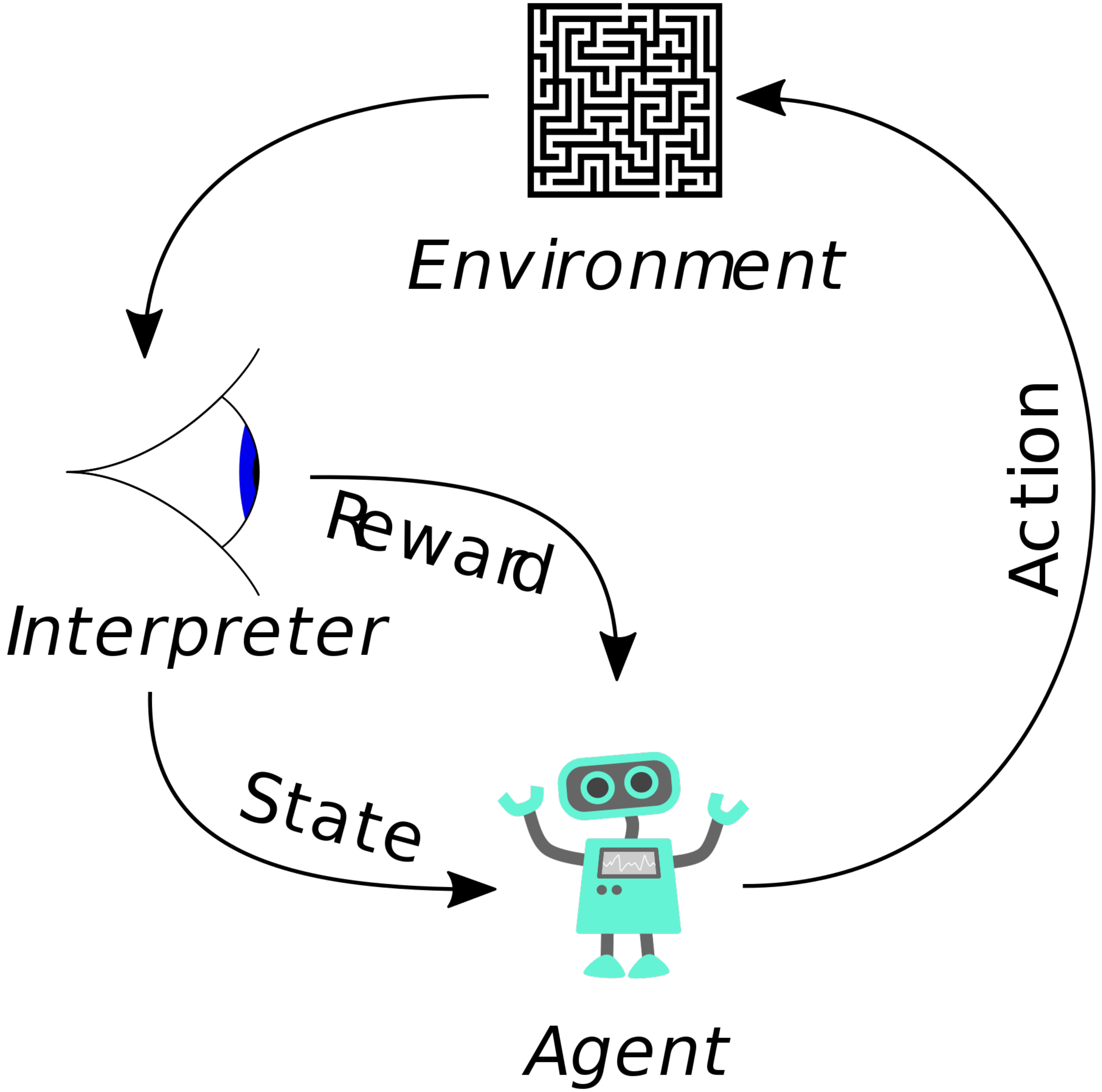
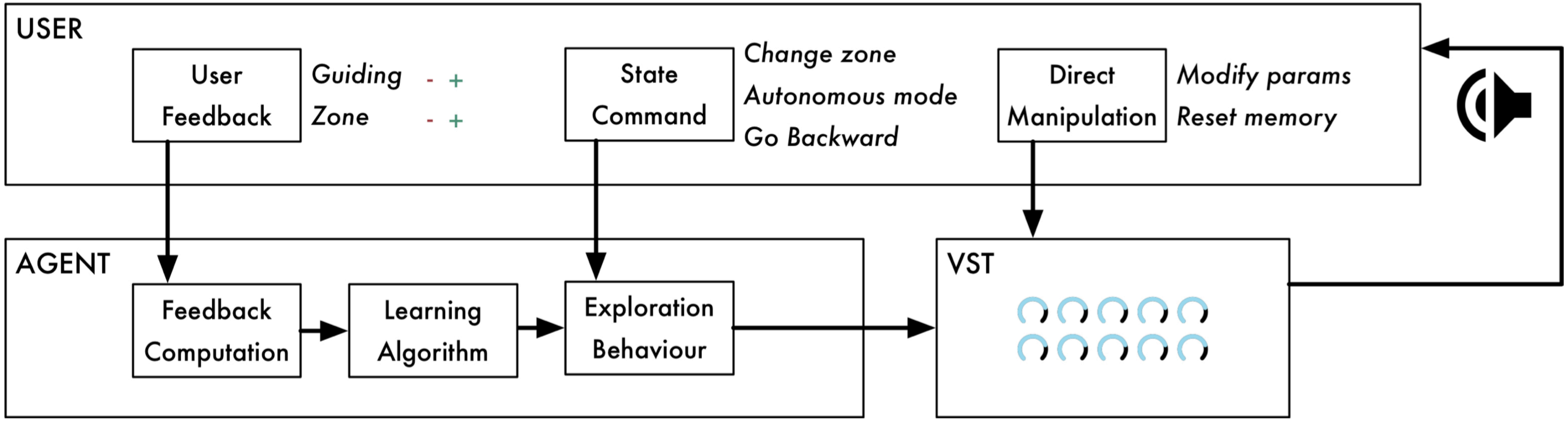


Image: Megajuice on Wikipedia

Reinforcement Learning

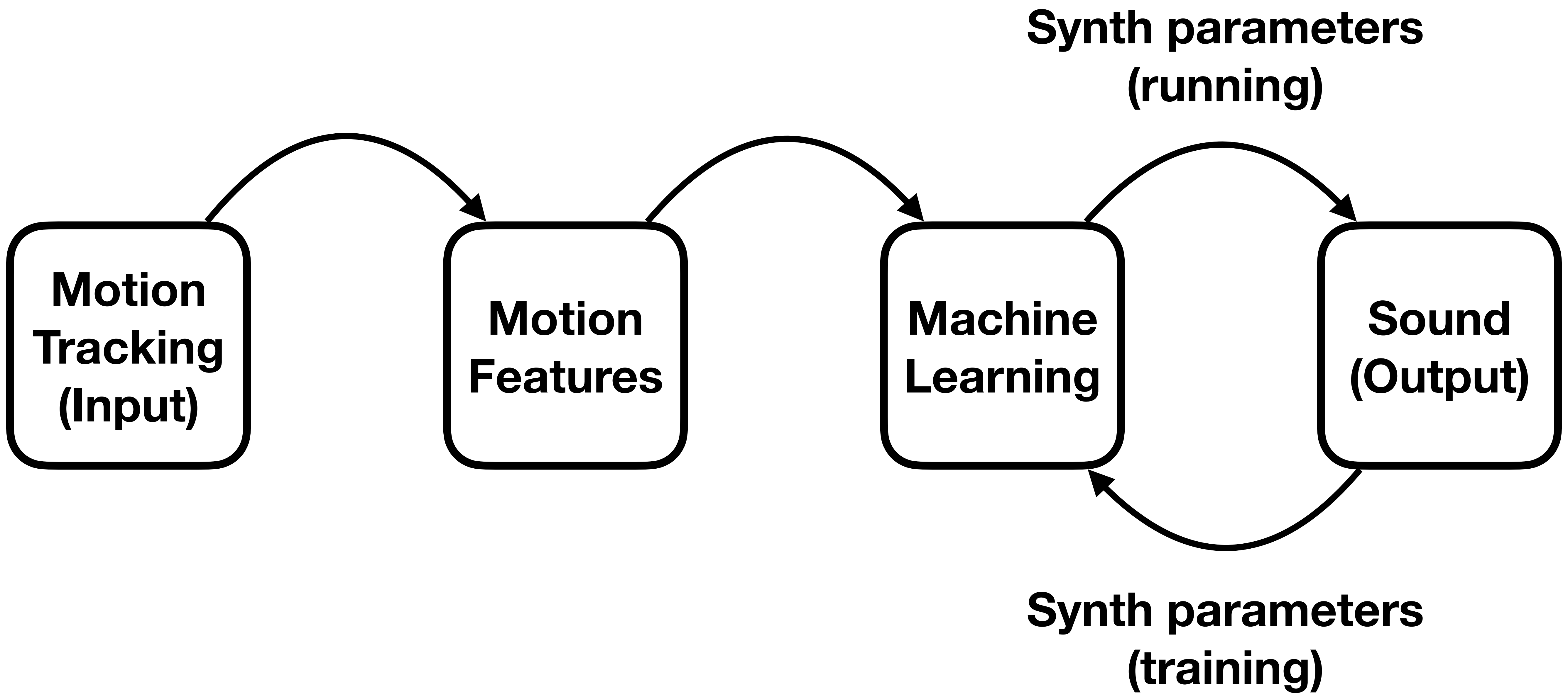
a sound design application



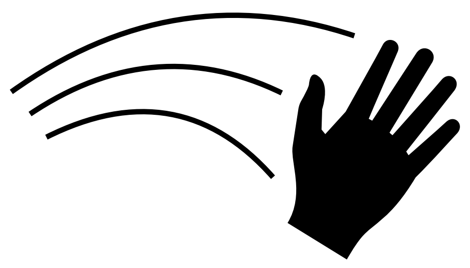
Scurto, H., Van Kerrebroeck, B., Caramiaux, B., & Bevilacqua, F. (2019). Designing Deep Reinforcement Learning for Human Parameter Exploration. ArXiv Preprint. Retrieved from https://arxiv.org/pdf/1907.00824.pdf

Interactive Machine Learning Model

What: control gestures



What: sounds

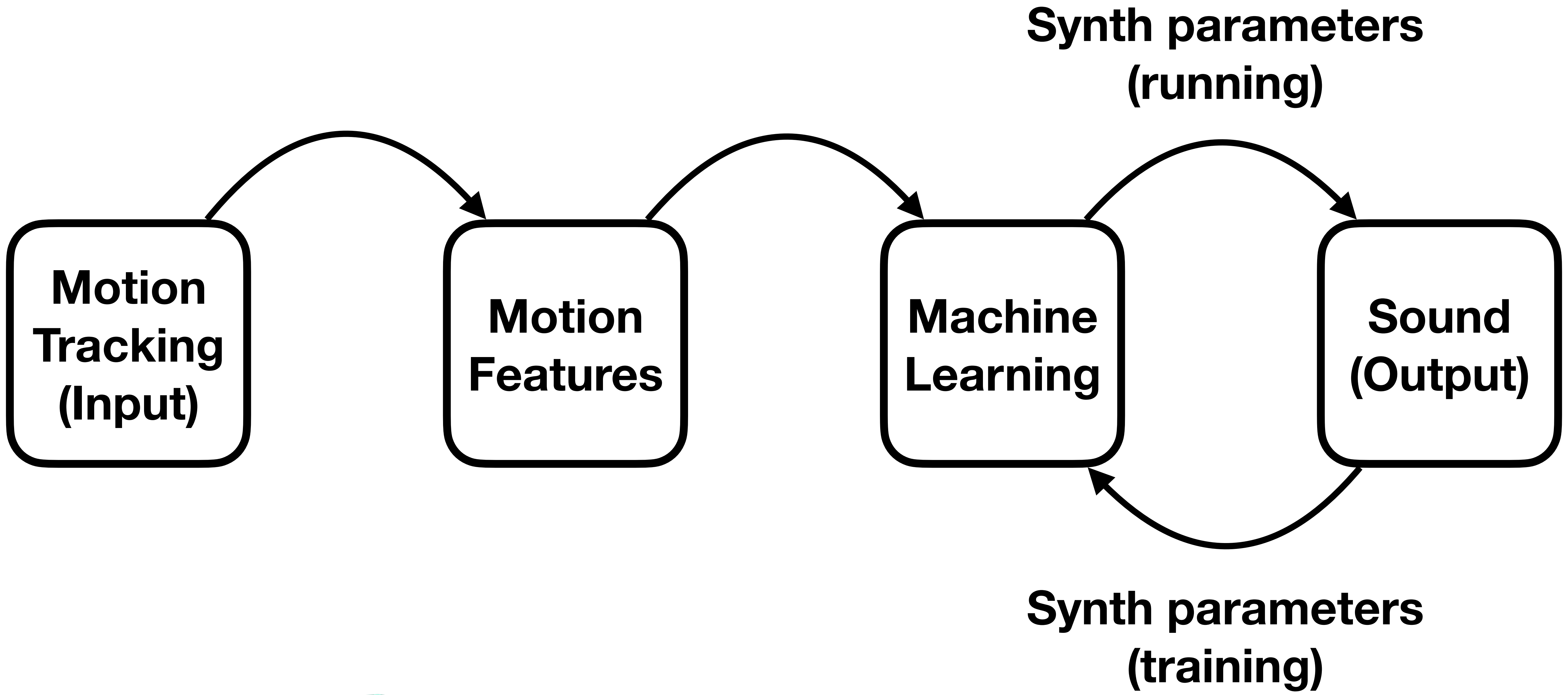


Examples

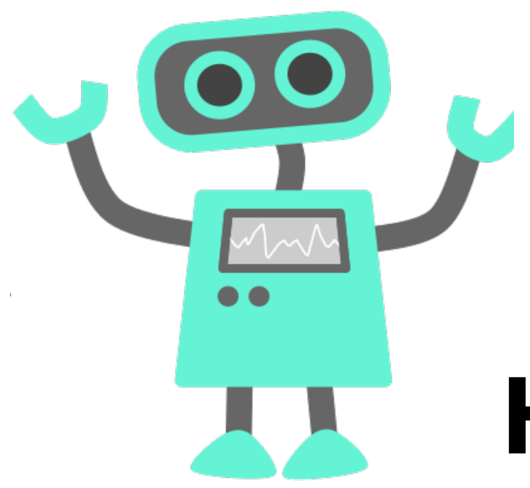
How these are related

Assisted Interactive Machine Learning Model

What: control gestures



What: sounds



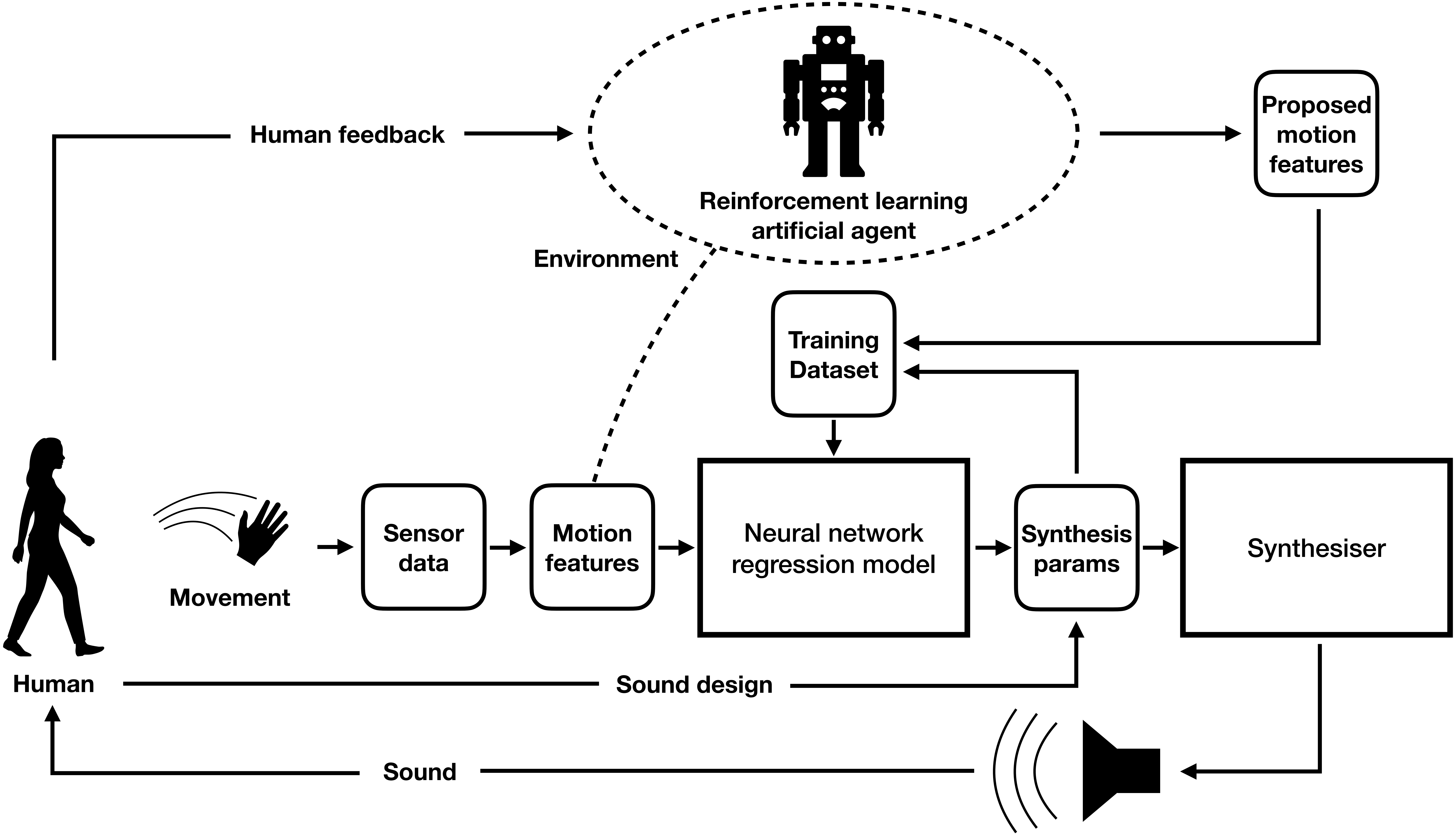
How these are related

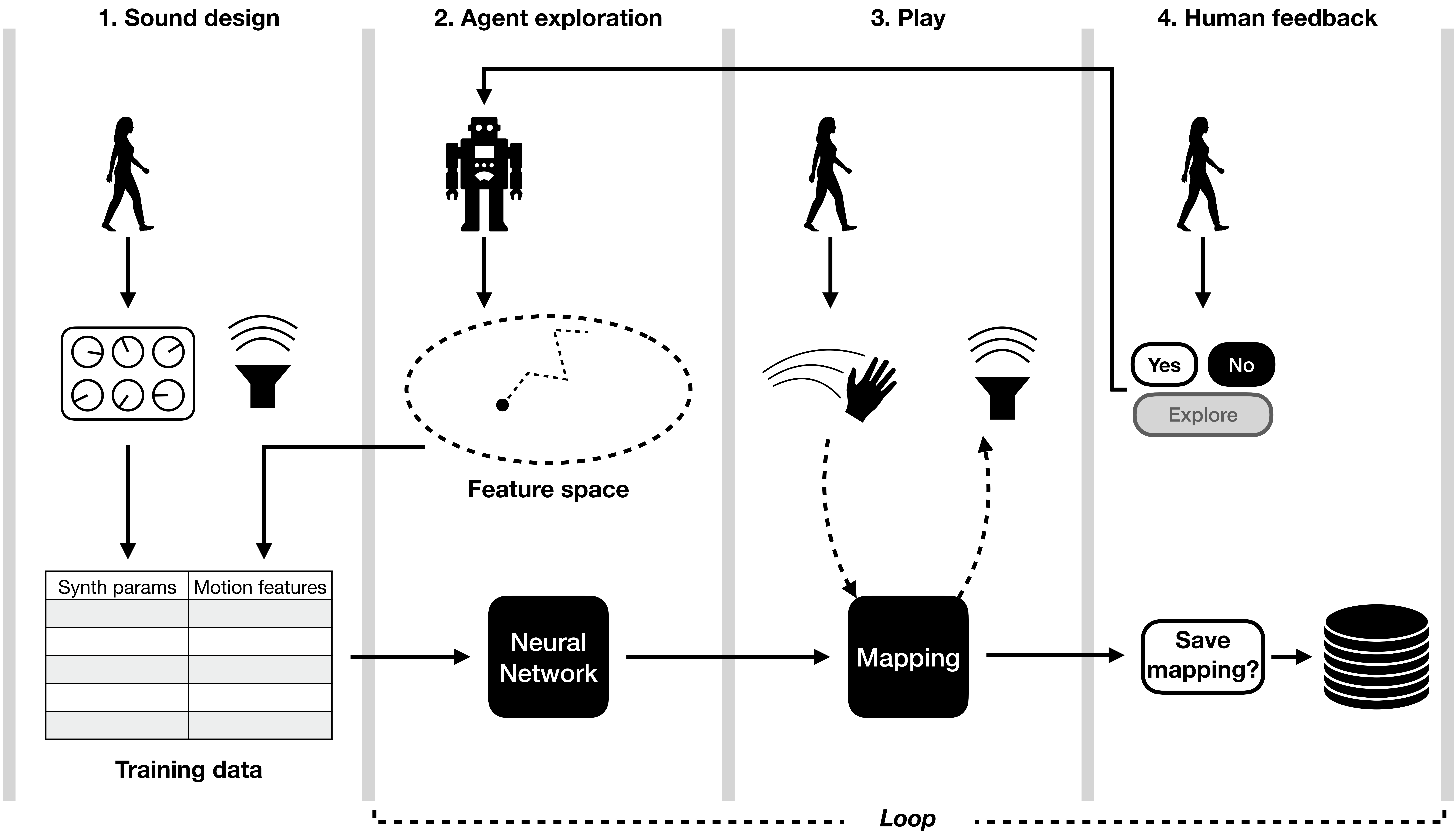
Agent

Interactive Machine Learning

Artificial Agent Exploration

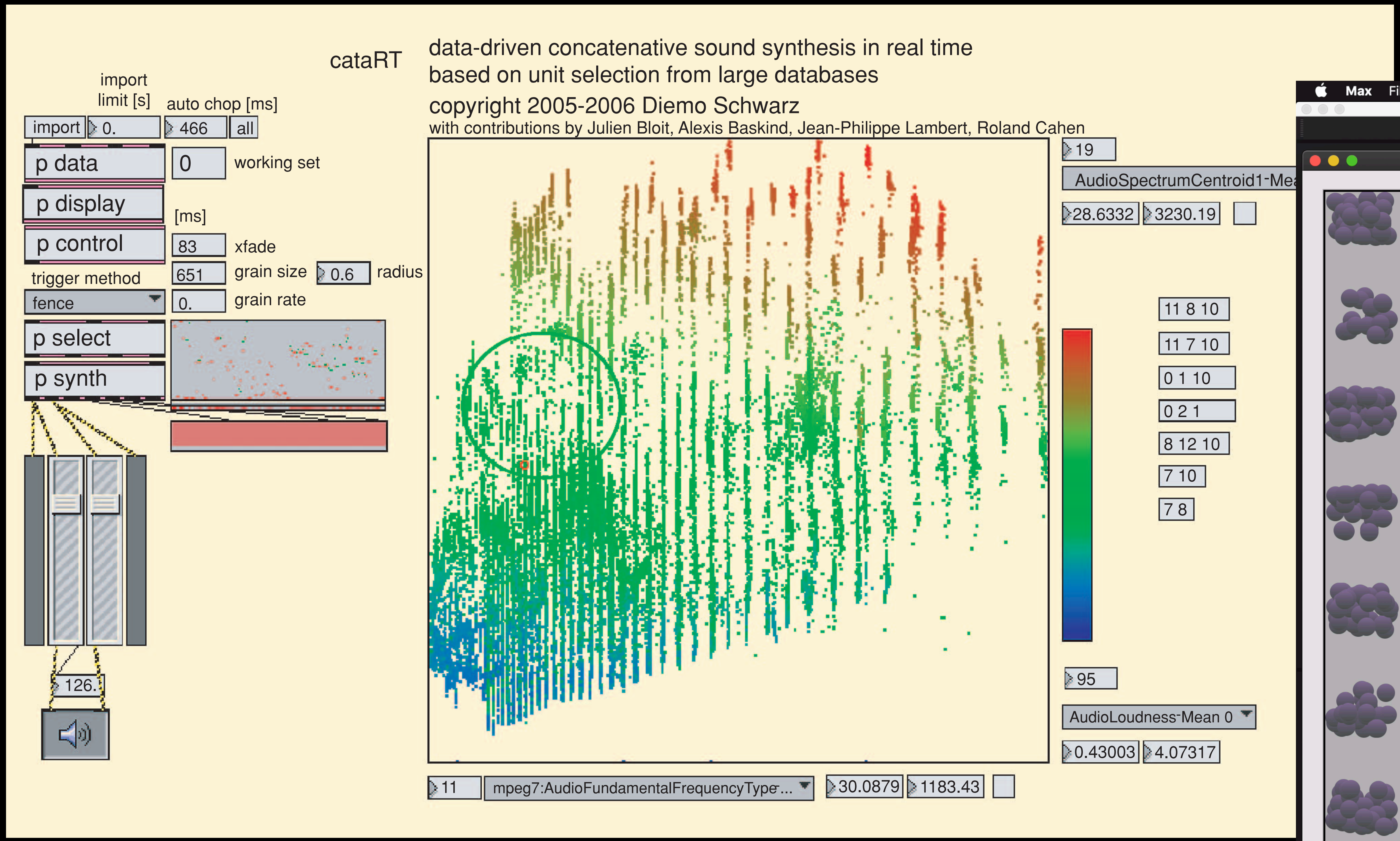
Assisted Interactive Machine Learning

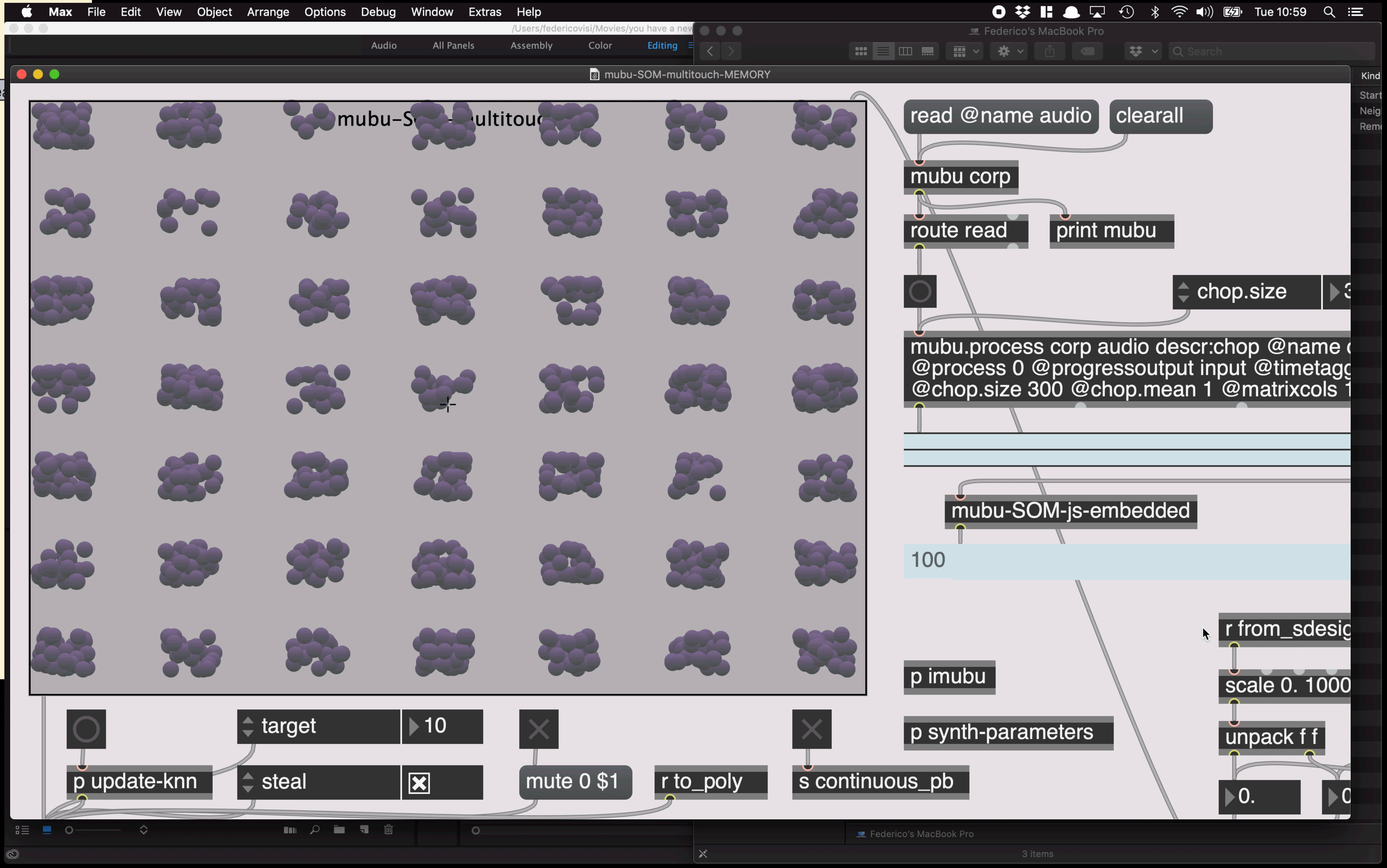

Human feedback
Reinforcement learning
artificial agent
Proposed motion features
Environment
Training Dataset
Human
Movement
Sensor data
Motion features
Neural network regression model
Synthesis params
Synthesiser
Sound design
Sound



1. Sound design
2. Agent exploration
3. Play
4. Human feedback
Feature space
Yes
No
Explore
Synth params
Motion features
Training data
Neural Network
Mapping
Save mapping?
Loop

Demo with corpus-based concatenative synthesis





Schwarz, D. (2007). Corpus-based concatenative synthesis. IEEE signal processing magazine, 24(2), 92-104.

Margraf, J. (2019). Self-Organizing Maps for Sound Corpus Organization. Masters Thesis. Technische Universität Berlin.

Using AIML to explore sonic memories

References

cutt.ly/visi-mapping-symposium

Visi, F. G., & Tanaka, A. (2021). **Interactive Machine Learning of Musical Gesture.** In E. R. Miranda (Ed.), *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Springer Nature, forthcoming.

Visi, F. G., & Tanaka, A. (2020). **Towards Assisted Interactive Machine Learning: Exploring Gesture-Sound Mappings Using Reinforcement Learning.** In *ICLI 2020 – the Fifth International Conference on Live Interfaces*.

Interactive Machine Learning of Musical Gesture

Federico Ghelli Visi* and Atau Tanaka **

GEMM))) Gesture Embodiment and Machines in Music
School of Music in Piteå, Luleå University of Technology, Sweden

EAVI – Embodied Audiovisual Interaction
Goldsmiths, University of London, UK

mail@federicovisi.com,
a.tanaka@gold.ac.uk

Abstract. This chapter presents an overview of interactive machine learning techniques applied to
tures. We go through the main c
analysing, and applying machi
gestures with the purpose of p
We discuss how different algori
tasks, including interacting wit
ploring interaction possibilities
conclude the chapter with a des
were employed by the authors fo
thus outlining the implications t
musical practice.

Keywords: machine learning,
analysis, expressive movement,
reinforcement learning, music.

Proceedings of the 5th International Conference on Live Interfaces, NTNU, Trondheim, Norway 2020

Towards Assisted Interactive Machine Learning: Exploring Gesture-Sound Mappings Using Reinforcement Learning

Federico Ghelli Visi[1, 2, *] and Atau Tanaka[2, **]

[1]GEMM))) Gesture Embodiment and Machines in Music
School of Music in Piteå, Luleå University of Technology, Sweden

[2]EAVI – Embodied Audiovisual Interaction
Goldsmiths, University of London, UK

[*]mail@federicovisi.com
[**]a.tanaka@gold.ac.uk

Abstract.

We present a sonic interaction design approach that makes use of deep reinforcement learning to explore many mapping possibilities between input sensor data streams and sound synthesis parameters. The user can give feedback to an artificial agent about the mappings proposed by the latter while playing the synthesiser and trying the new mappings on the fly. The design approach we adopted is inspired by the ideas established by the interactive machine learning paradigm, as well as by the use of artificial agents in computer music for exploring complex parameter spaces. We refer to this interaction design approach as Assisted Interactive Machine Learning (AIML). We describe the architecture of an AIML system prototype, a typical workflow for interacting with the agent and obtain gesture-sound mappings. We then present feedback data collected during a demonstration and discuss perspectives for developing the AIML paradigm further, pointing out current limitations. In light of the feedback obtained and the considerations arisen following the use of the system in a multimedia performance piece, we suggest that the implementation and evaluation of new features should take into consideration established creative workflows and take place close to actual artistic practice.

Keywords. Gestural Interaction, Interactive Machine Learning, Reinforcement Learning, Artificial Agents, Sonic Interaction Design

Reflections

- In actual artistic practice: AIML as something creatively useful, able to surprise and help breaking creative deadlocks
- Helpful, inspiring, not substituting people's creativity
- Interplay with established practices
- Supporting exploration of complexity, in parameter spaces and data corpora
- Learning: understanding how things works through co-exploring
- Demystifying AI/ML by using it actively to build something

Thank you.

Contact: mail@federicovisi.com

Referenced articles: cutt.ly/visi-mapping-symposium

Special thanks to my collaborators in these studies:
Stefan Östersjö, Robert Ek, Ulrik Röijezon, the GEMM))) cluster,
Atau Tanaka, Michael Zbyszyński, Jonas Margraf