# Self-Awareness in a Cyber-Physical Predictive Musical Interface

Charles P. Martin[1,2], Kyrre Glette[1,2], Tønnes F. Nygaard[1], and Jim Torresen[1,2]

[1]Department of Informatics and [2]RITMO Centre, University of Oslo, Norway
charlepm@ifi.uio.no

### Abstract

We introduce a new self-contained and self-aware interface for musical expression where a recurrent neural network (RNN) is integrated into a physical instrument design. The system includes levers for physical input and output, a speaker system, and an integrated single-board computer. The RNN serves as an internal model of the user's physical input, and predictions can replace or complement direct sonic and physical control by the user. We explore this device in terms of different interaction configurations and learned models according to frameworks of self-aware cyber-physical systems.

## Introduction

In this paper, we introduce a prototype new interface for musical expression (NIME) integrating a recurrent neural network (RNN) into a physical musical interface. The interface includes levers for physical input and output, a speaker system, and an integrated Raspberry Pi for sound synthesis and RNN prediction. The RNN serves as an internal model of the user's physical input, this model is able to predict the user's next input based on their past inputs. The presence of a self-aware internal model, physical interfaces, and the potential for network communication makes this a cyber-physical system (CPS) for creative interaction.

We investigate this instrument from a computational self-awareness perspective. We use our system's flexible architecture to experiment with various musical and interaction scenarios, and aim to establish that our system possesses stimulus awareness, interaction awareness, and time awareness. Our system can be configured to generate an extra layer of musical performance on top of the user's, as a filter for the user's input, or in turn-taking interactions. We define such real-time loops between a human's creative input and machine-learning predictions as predictive creative interaction. We present multiple predictive creative interaction schemes with different configurations of internal model and synthesis mapping, and explore how these are connected with aspects of self-awareness.

Embedded and self-contained instruments are important current topics in digital musical instrument design (Moro et al., 2016); however, these instruments usually do not include self-awareness. On the other hand, musical AI is often focussed on composition using high-level symbolic representations (e.g., Sturm and Ben-Tal, 2017), and not the interactive or embodied factors (Leman et al., 2018) of music perception and creation. In this work, an embedded instrument design is combined with a novel, embodied approach to musical AI. By combining self-awareness with musical interaction, this instrument can be considered both as an interface for human expression, and an artificial-life based artwork. Musical instruments that, like our system, include artificial life and self-awareness attributes, could lead to new kinds of musical expression, where performers engage in interactive embodied collaboration with their instruments.

Figure 1: Our prototype interface; the left lever allows user input, and the right lever represents RNN output. A demo can be viewed at `https://youtu.be/0XQ0hRCsbhU`

## System Design

Our interface (Figure 1) is a self-contained cyber-physical system. The 3D-printed enclosure includes a Raspberry Pi, a touch-screen and lever for input, a speaker and servo-controlled lever for physical output, and a USB battery. Self-awareness is provided by an internal model of the user input: a sequence of real-valued potentiometer positions, along with a time-delta value. To model this data, we use a 2D mixture density RNN that predicts the position, and the time, of the next user input. This neural network de-

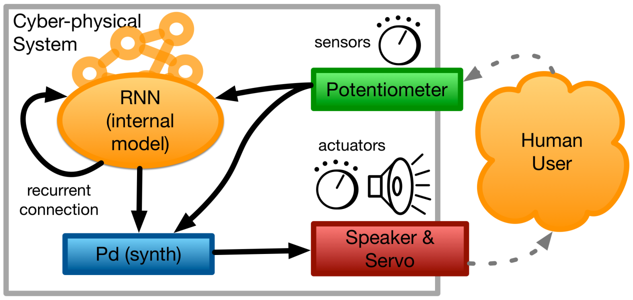


Figure 2: System diagram of our interface; the internal connections can be configured in several ways.

sign was introduced by Bishop (1994), but has only recently been applied to musical interaction data (Martin and Torresen, 2018). Various trained models can be used with this network based on either real-world or synthetic training data.

The system is implemented in Python using TensorFlow[1]. This software manages input and output from the microcontroller, predictions from the MDRNN. Sound synthesis is performed in Pure Data which is controlled via open sound control (OSC) from the Python application. A system diagram is shown in Figure 2.

## Interaction and Self-Awareness

We aim to establish that the presented system possesses stimulus awareness, interaction awareness, and time awareness (Lewis et al., 2015). We explore how the interactive configuration of our system can be adjusted by connecting sensors, actuators, and learned models in different ways. In our system, the user can always use their lever to control sound, but connections to and from the RNN model are optional: the user's input can be driven through the RNN to condition its memory state, the RNN output can be used to drive synthesised sound and its own lever, and the RNNs output can be connected to its input for unconstrained predictions. Further, learned models (trained on real-world or synthesised data), and conditioned memory states, can be exchanged within our system and potentially between others. We aim to implement goal and meta-self-awareness by automating these interactive connections and model configurations. At present, we have explored three interaction scenarios:

**Call-and-Response.** The user controls their lever to trigger sound and condition the RNN memory; if there is no input for two seconds, the RNN continues from its memory state until the user interacts again.

**Harmony.** The user controls their lever to trigger sound and condition the RNN memory state; the RNNs output predictions are used to control its own lever and sound.

[1] `https://github.com/cpmpercussion/creative-mdns`

**Battle Royale.** Both the user and RNN are in control of separate sounds, the RNN continually generates control data regardless of the user's input.

The system's ability to sense the user's actions, as well as it's own actions (through the RNNs recurrent connections) gives stimulus awareness. The above modes of musical interaction show interaction awareness at a level above stimulus awareness. Finally, the RNN model provide temporal awareness due to its recurrent connections, LSTM memory cells, and ability to make precise temporal predictions.

## Conclusions and Future Work

In this work, we have contributed a novel, self-aware musical instrument, and shown how it satisfies several criteria of self-awareness. This system can be considered as an interface for human expression, or as an independent artificial musician that is able to sense, and collaborate with, humans. For future work, we plan to evaluate our system with performers in different musical scenarios. We also plan to share knowledge between groups of these systems which could be accomplished through stimulus sharing, sharing of interaction episodes, or sharing of independently learned RNN models. Such systems could be compelling as both artificial-life based musical artworks, and interfaces for enhancing and enabling human musical creativity.

**Acknowledgements** This work was supported by the Research Council of Norway through the EPEC project (#240862), and its Centres of Excellence scheme (#262762).

## References

Bishop, C. M. (1994). Mixture density networks. Technical Report NCRG/97/004, Neural Computing Research Group, Aston University.

Leman, M., Maes, P.-J., Nijs, L., and Van Dyck, E. (2018). What is embodied music cognition? In Bader, R., editor, *Springer Handbook of Systematic Musicology*, pages 747–760. Springer Berlin Heidelberg, Berlin, Heidelberg.

Lewis, P. R., Chandra, A., Faniyi, F., Glette, K., Chen, T., Bahsoon, R., Torresen, J., and Yao, X. (2015). Architectural aspects of self-aware and self-expressive systems: From psychology to engineering. *Computer*, 48(8):62–70.

Martin, C. P. and Torresen, J. (2018). Robojam: A musical mixture density network for collaborative touchscreen interaction. In Liapis, A., Romero Cardalda, J. J., and Ekárt, A., editors, *Computational Intelligence in Music, Sound, Art and Design*, pages 161–176, Cham. Springer International Publishing.

Moro, G., Bin, A., Jack, R. H., Heinrichs, C., McPherson, A. P., et al. (2016). Making high-performance embedded instruments with bela and pure data. In *International Conference on Live Interfaces*. University of Sussex.

Sturm, B. L. and Ben-Tal, O. (2017). Taking the models back to music practice: Evaluating generative transcription models built using deep learning. *Journal of Creative Music Systems*, 2(1).