Incisions: Tangible Latent Space Exploration with Three Sound Balls

Matthias Jung
Institute for Music, Science and Engineering
King Mongkut's Institute of Technology Lat Krabang
Bangkok (Thailand)
matthias.ju@kmitl.ac.th

ABSTRACT

This system proposes an interactive, tactile approach to collaboratively exploring machine learning models in real-time. The current work-in-progress design connects three handheld spherical devices (sound balls) to three machine learning models. Equipped with pressure sensors and gyroscopes, the sound balls send readings from an ESP32 as OSC over Wi-Fi to a Max/MSP patch that hosts the model playback. The patch employs a combination of open-source and self-trained models, which are mixed into a master output audible through headphones by the three sound ball players, who collaboratively explore the models within a latent-dimension setup.

Keywords latent space exploration, RAVE, sound balls, tangible multi-user instruments, co-creation, accessible musical instruments

1. INTRODUCTION

Spherical handheld interfaces for musical expression (referred to sound balls here below) have been used as both standalone musical instruments, as well as interfaces for participatory works without integrated playback. They have been proven to be accessible and playful devices, and are well suited for interactive or collaborative creative scenarios. With recent micro-controllers such as the ESP32 it has become easy to send sensor readings over WiFi and fit everything into small, handheld balls with little weight. At the same time, there have been advances made in real-time use of machine learning models, and those models have been integrated in commonly used audio production and performance software during recent years. Here, RAVE (real-time variational auto-encoder) with its nn∼ Max/MSP-object is a prominent example, allowing to explore timbre spaces of audio datasets via a set of latent parameters. The suggested system makes use of both developments and suggests to combine the affordances of tactile interfaces with RAVE. In addition, it uses three balls simultaneously to control the latent parameter space among three RAVE models on different levels.

2. BACKGROUND AND RELATED WORK

2.1 Spherical Handheld Interfaces for Musical Expression

Early examples for sound ball designs include Weinberg et al.'s "Embroidered Musical Ball" [1, 283], and Jensenius et al.'s "Music Balls" [2], both designed to include musical novices as users of those accessible small musical instruments. Many of those instruments explored different flex and force-sensing resistors or conductive thread as sensors, feeding the sensor readings in desktop computers with a cable-bound interface, then mapped to different sounds via MIDI.

The Bluetooth Radio Ball Interface (BRBI) introduced a wireless handheld foam ball integrating a 3-axis accelerometer, transmitting sensor data via Bluetooth as OSC messages for sound generation [3]. While not strictly spherical, Bowen's Soundstone similarly employs a Bluetooth-enabled accelerometer-based interface for gestural input to Max/MSP, extending interaction through integrated visual and haptic feedback using a tri-color LED and a vibrator [4]. Extending the scale and bodily engagement of such devices, the Balance Ball enables full-body interaction through rolling, lifting, and riding gestures. It incorporates multiple sensors, including accelerometer, gyroscope, photodiode, and LEDs, and maps motion data to velocity, pitch, and tempo MIDI messages for musical control [5].

Collaborative ball-based instruments emphasize shared control and accessibility. The StringBall is designed for two performers connected via opposing strings passing through the ball, allowing lateral motion, lifting, and shaking gestures to produce chaotic sound textures [6]. Its simplicity targets inclusive use by non-musicians and children. In contrast, the commercial Oddball [1] supports complex rhythmic interaction through learned bouncing gestures mapped via a smartphone application. While offering advanced musical capabilities, its cost limits scalability for educational contexts and participatory performance.

2.2 Latent Space Navigation in RAVE

Real-time audio synthesis based on prior training of machine learning models with audio datasets have been made accessible in the last years with Somax [7], DDSP [8], or RAVE [9], to name some of the widely used examples. The model used for both training and real-time synthesis in Max/MSP in this work is the latter: Realtime Audio Vari-

[1] Oddball https://oddballism.com/en-eu

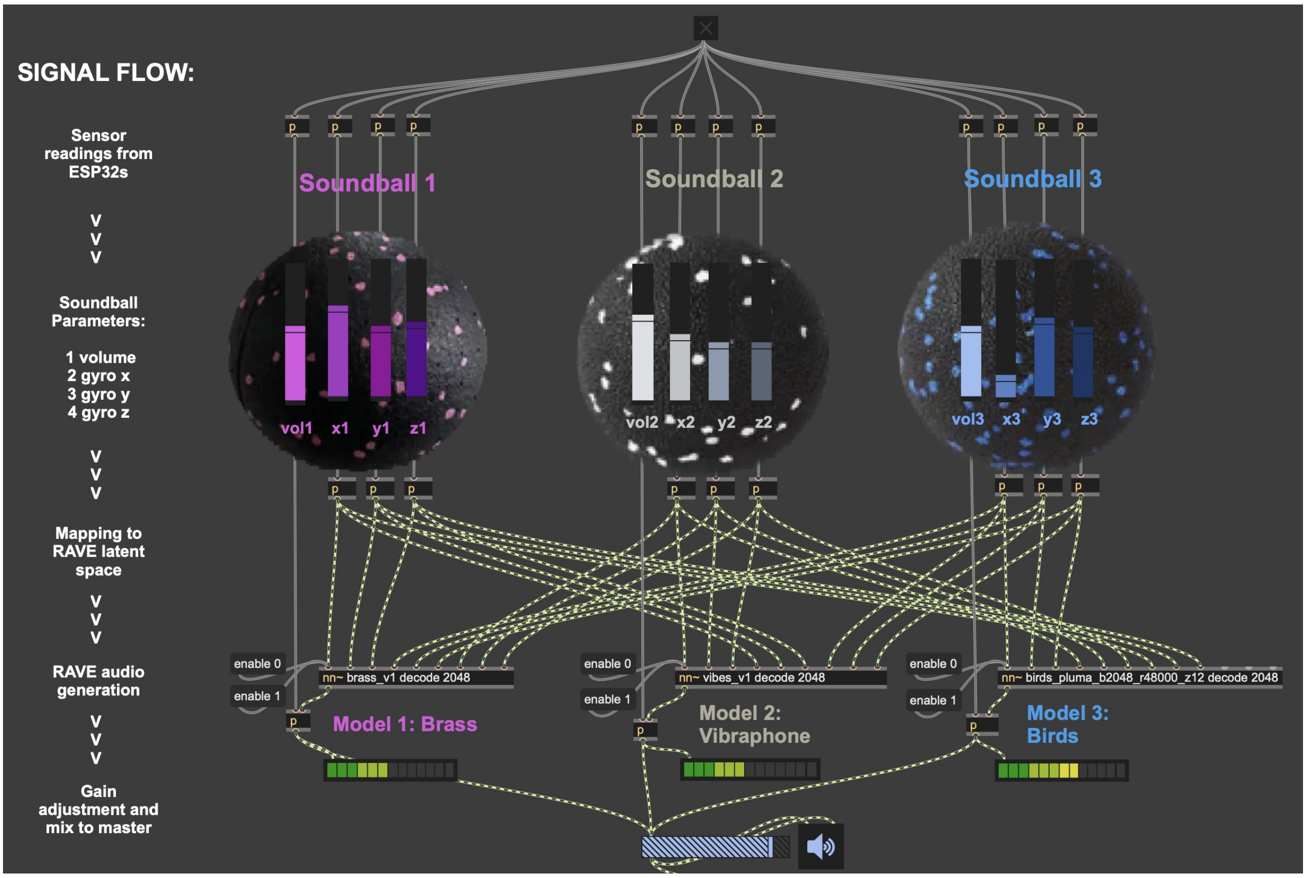

Figure 1. Max/MSP patch for Incisions.

ational autoEncoder (RAVE) [2] . Those tools make it possible to explore creative novelty by interacting with machine learning models trained in advance, then used for musical performance in real-time [10]. RAVE's decode function in the nn∼ object for Max/MSP also allows synthesis without audio input, i.e. the descriptor-based generation via latent space parameters only [11].

3. SYSTEM DESIGN AND ARCHITECTURE

The title of this paper points to the idea of "cutting through the timbre space," very much like an incision. The system uses the feature of the nn∼ object by mapping the sensor readings of the three sound balls to three nn∼ objects using the decode function for descriptor-based generation based on those variables. The signal flow is shown in figure 1 and can be described by the following three stages:

1. Sending and receiving sensor readings: The four sensor readings are being sent from each of the three sound balls' ESP32s to Max/MSP over Wi-Fi [3] .

2. Mapping sensor readings to control parameters in Max/MSP: The pressure sensor reading is mapped to the volume control of each of the three RAVE models. The x,y,z positional readings of the sound balls are mapped to the first three parameters of each or the adjunct model of each sound ball. The gyroscope readings of each sound ball are also mapped to the inlets 4-6, and 7-9, of the other two models respectively. The readings therefore are not only used for one model, but also translated to both other model's latent variables. The fact that the first variables have more influence on the generated audio output of each model takes care of the fact that the main agency of each player is still remains with "their" sound ball.

3. Mixing and playback: All model output is mixed together into a master output based on the pressure readings and played back via an external interface with 3 headphones so that all players hear the same master audio output.

The appended video shows a simulation of the changing values of the sound ball based on random up/down movements, as would be the case when three human players are using the system. However, it has to be noted that humans will move in a very different fashion and especially the volumes are supposed to be controlled in close interaction, i.e. people might listen to each others models, try to hold still for a while, or take turns in their explorations of the timbre spaces. At the time of writing, we also explore the introduction of an additional layer of musical parameters (such as pitch, delay, drive, reverb, etc.) mapped to the latent space parameters, which is showing some promising results for future designs.

[2] https://github.com/acids-ircam/RAVE

[3] This has been the setup for previous participatory work with the sound balls and works seamlessly for three sound balls.

3.1 Sound Ball Interface

As described in section 2, the sound ball follows a tradition of wireless, handheld sensor balls. The author has experimented with different designs during the last year and now built the sound balls used for this work from massage balls, that are made of denser foam material. This makes it easy to control the pressure sensor more accurately in contrast to balls with softer material. The limitation to just one pressure sensor was found to work best when controlling gyroscope parameters, and the mapping of the pressure sensor to the individual volume is easy to understand for novices to regulate ”their” contribution to the master mix.

Figure 2 shows a photo of the sound ball to be used for the interactive demo, while figure 3 shows a design sketch including its components. Those are as follows:

1. A massage ball made of hard foam, carved out from the inside to about 1cm and re-assembled after components were built in.

2. An on/off switch and a 3.7V 2000mAh lithium battery pack to power the ball for around 2-3 hours (usually enough for a concert lengths or exhibition period).

3. A loading socket (USB-C) and TP4056 loading circuit.

4. An MCU with Wi-Fi (ESP32) to act as a Websocket client for the Max/MSP-bound server.

5. A pressure sensor (FSR-402) and a gyroscope (MPU-6050).

Figure 2. Current prototype of the Sound Ball.

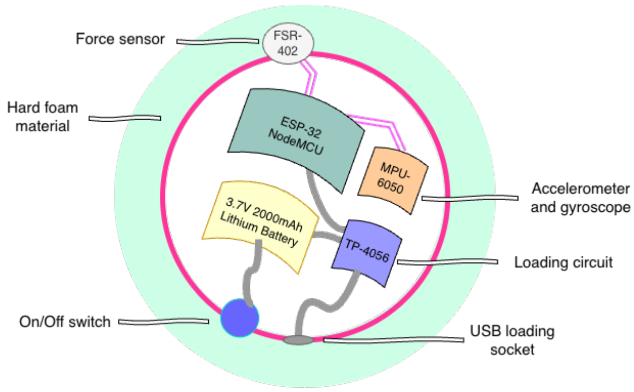

Figure 3. Design of the current prototype of the Sound Ball with components.

3.2 Datasets and Training

For testing the system design, the following three RAVE models were used inside a Max/MSP patch with the nn∼ object using its decode function.

- **Model 1** For the first model, the Philharmonia brass recordings were used [4], specifically the instrument categories trumpet, trombone, and tuba. This lead up to a total length of the audio dataset of 15min and 21 sec. The RAVE v2 model was trained at 44.1k sample rate, a compression ratio of 512, 57.6 M total parameters and 9 latent dimensions. This number was chosen due to the fact that each of the gyroscopes from the sound balls are sending 3 values, i.e. 9 values to use for the latent parameters in total. Training was concluded after 32000 steps and the model size is 230.347 MB.

- **Model 2** The second model used for the experiment was generated studio recordings of a Musser vibraphone, recorded and played by the author recently. The total length of that dataset was only 3min and 7sec long and used similar parameters as model 1. However, training was concluded already after 16000 steps given the small dataset.

- **Model 3** As a third model, a RAVE model trained on bird recordings curated by Giacomo Lepri, from the RAVE collection by the Intelligent Instruments Lab [12] was used [5] .

4. CONCLUSIONS AND FURTHER WORK

A tangible interface for exploring RAVE models collaboratively in real-time has been suggested for investigation and a prototype has been built based on the previously proposed sound ball instrument. The pressure sensor is used to control the individual volume of each sound ball, while the x,y,z readings of the gyroscope sensor are being used to control the latent space parameters of three RAVE models. Results are preliminary and more models will be generated for further testing the system design. Further work should

[4] The licence for this dataset is CC 3.0 and the data was acquired from https://philharmonia.co.uk/resources/sound-samples.

[5] CC-BY-NC-4.0 licensed model ”birds_pluma_b2048_r48000_z12.ts” from https://huggingface.co/Intelligent-Instruments-Lab/rave-models.

also include different setups for controlling the other players' latent space parameters, after the demo has been tested with both musical amateurs and professionals, as well as discussed with experts.

Acknowledgments

This work was kindly supported by the King Mongkut's Institute of Technology Ladkrabang Research Fund (Research Start-up Grant for New Faculty Members, grant number KREF186907). The author thanks both the project mentor Associate Prof. Dr. Sorasak Danworaphong of the Institute of Music, Science and Engineering, King Mongkut's Institute of Technology Ladkrabang, as well as the co-principle project investigator Prof. Dr. Stefan Weinzierl of the Audio Communication Group, Technische Universität Berlin.

5. REFERENCES

[1] G. Weinberg, M. Orth, and P. Russo, "The Embroidered Musical Ball: A Squeezable Instrument for Expressive Performance," *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, pp. 283–284, 2000. [Online]. Available: https://doi.org/10.1145/633292.633457

[2] A. R. Jensenius and A. Voldsund, "The Music Ball Project: Concept, Design, Development, Performance," *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2012. [Online]. Available: http://www.nime.org/proceedings/2012/nime2012_161.pdf

[3] W. S. Yeo, "The Bluetooth Radio Ball Interface (BRBI): A wireless Interface for Music/Sound Control and Motion Sonification," *Proceedings of the International Computer Music Conference*, 2006.

[4] A. Bowen, "Soundstone: A 3-D Wireless Music Controller," *Proceedings of the 2005 Conference on New Interfaces for Musical Expression*, pp. 268–269, 2005.

[5] T. Yamaguchi, T. Kobayashi, and S. Hashimoto, "Balance Ball Interface for Performing Arts," *Human Interface and the Management of Information. Interacting with Information: Symposium on Human Interface 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part II*, pp. 440–445, 2011.

[6] W. S. Yeo, J. W. Yoon, and H. Y. Cho, "The Stringball: A Ball Interface With Strings for Collaborative Performance," in *Proceedings of the International Computer Music Conference*, 2007.

[7] J. Borg, "Somax 2: A Real-time Framework for Human-Machine Improvisation," *STMS IRCAM & Aalborg University, Internal Student Report*, 2019.

[8] J. Engel, L. Hantrakul, C. Gu, and A. Roberts, "DDSP: Differentiable digital signal processing," *Proceedings of the Eighth International Conference on Learning Representations*, 2020. [Online]. Available: https://arxiv.org/pdf/2001.04643

[9] A. Caillon and P. Esling, "RAVE: A Variational Autoencoder for Fast and High-Quality Neural Audio Synthesis," in *Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx)*, 2021. [Online]. Available: https://arxiv.org/abs/2111.05011

[10] A. Lunt and S. Trump, "Latent Space Explorer," *Proceedings of the International Conference on AI Music Creativity (AIMC) 2023*, August 29 2023, https://aimc2023.pubpub.org/pub/zgc5j7ha.

[11] A. Caillon, A. Bitton, B. Gatinet, and P. Esling, "Timbre latent space: exploration and creative aspects," 2020. [Online]. Available: https://arxiv.org/abs/2008.01370

[12] Intelligent Instruments Lab, "rave-models (Revision ad15daf)," 2023. [Online]. Available: https://huggingface.co/Intelligent-Instruments-Lab/rave-models