INCLUSIVE PERFORMATIVE ECOSYSTEMS: AI, BCI, AND BIOFEEDBACK FOR ACCESSIBLE MUSICAL INTERACTION

Sara MONTAGNI (montagnisara@gmail.com) [1]

[1] **Luigi Boccherini Conservatory**, Lucca, Italy

ABSTRACT

This paper presents a hybrid interaction system engineered for inclusive musical performance, investigating how physiological signals and computer vision can enable real-time, hands-free control of musical structures. The architecture integrates two complementary modalities: EEG signals acquired via an OpenBCI headset and facial gesture tracking using MediaPipe FaceMesh. Raw EEG data are processed through a custom Python bridge that extracts spectral and statistical features to train machine learning models within Wekinator for classification and regression, with resulting parameters mapped via Open Sound Control (OSC) to generative processes in Ableton Live. In parallel, a webcam-based interface detects head movements to navigate a circular graphical UI, allowing discrete selection of root notes, scales, arpeggiator patterns, and tempo divisions through dwell-time interaction and specific facial gestures. By combining continuous neural control with discrete facial selection, the system enhances interaction stability and expressive capabilities, offering a novel approach to neuroadaptive instruments, particularly for users with limited motor abilities. The paper delineates the system architecture, interaction design, and signal-processing strategies. Preliminary single-participant tests demonstrate system feasibility, confirming a real-time total system latency below 10 ms and validating the viability of this multi-modal approach.

1. INTRODUCTION

Research on brain–computer interfaces (BCIs) for musical interaction has expanded significantly over the past two decades, exploring ways to translate neural activity into expressive control of sound synthesis and generative musical processes. Early pioneering experiments in this domain include Alvin Lucier's *Music for Solo Performer* (1965) and the systematic investigations collected by Rosenboom [1], which established the foundations of biofeedback-driven musical performance. Subsequent research by Miranda and Brouse [2] formalised the notion of Brain-Computer Music Interface (BCMI), demonstrating how EEG-derived features could be mapped to synthesis parameters in real time. More recent research integrates machine learning techniques to interpret physiological signals and recognise patterns associated with cognitive states such as attention or relaxation [3]. At the same time, computer vision technologies have enabled new forms of gesture-based interaction. Facial landmark detection systems such as MediaPipe FaceMesh allow real-time tracking of facial geometry using standard webcams, enabling the detection of eye blinks, facial gestures and head movements without specialised hardware. Webcam-based tracking has also been investigated as a low-cost alternative to dedicated eye-tracking devices for accessible gaze-like interaction [4].

The choice of EEG as the primary physiological signal is motivated by two considerations. First, the involuntary nature of neural modulations: rather than treating brain activity as a deliberate command channel, the system exploits the slow, partly involuntary dynamics of spectral band power to drive continuous sonic modulation, so that the performer's internal state becomes an expressive material in itself. Second, EEG provides long-term, low-frequency descriptors (e.g. theta/alpha balance) that are well suited to shaping musical processes over extended time scales. Alternative and complementary physiological signals—in particular a respiratory belt and ECG—are considered as future extensions of the sensing layer (Section 6).

The conceptual framework of this research draws on Waters' notion of performative ecosystems [5], which reconceives the performer–instrument–audience triad as a dynamic, non-hierarchical network in which biosignal interfaces function as active agents rather than mere controllers. Despite these developments, relatively few musical systems integrate neural signals, machine learning and computer vision within a unified interaction framework. This paper contributes: (i) a hybrid interaction architecture combining EEG-based neurofeedback with facial gesture control; (ii) a real-time bridge connecting OpenBCI, Wekinator and Ableton Live through the OSC protocol; (iii) a hands-free musical interface designed for inclusive performance scenarios.

1.1 Accessible Digital Musical Instruments

Since the system is intended to be usable by performers with limited motor abilities, it is situated within the field of Accessible Digital Musical Instruments (ADMIs), surveyed by Frid [6], whose review highlights the scarcity of interfaces designed for hands-free and inclusive interaction. Within this field, gaze- and face-based control has been proposed as a viable strategy for performers who cannot rely on manual gestures; Davanzo and Avanzini [7] dis-

Copyright: © 2026. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

cuss design concepts for gaze-based digital musical instruments and the trade-offs between selection accuracy and temporal responsiveness. The present work positions nose-pointer and blink-based control as a low-cost, webcam-only alternative within this design space.

2. SYSTEM ARCHITECTURE

The proposed system integrates neural sensing, machine learning and computer vision within a unified real-time architecture for interactive musical control. The architecture follows a modular design in which several processes communicate through the Open Sound Control (OSC) protocol [8]. The system includes three main components: EEG acquisition and signal processing, machine learning analysis and a face-tracking interaction interface. EEG signals are acquired using an OpenBCI headset [9] and transmitted via UDP from the OpenBCI graphical interface. A Python bridge receives these data streams, extracts signal features and sends them to Wekinator [10] for classification and regression. The outputs of the machine learning models are transmitted via OSC to Ableton Live, where they control musical parameters such as modulation depth and generative processes. In parallel, facial movements captured by a webcam are analysed using MediaPipe FaceMesh [11] and also transmitted via OSC. By combining neural and facial interaction, the system provides both continuous control derived from EEG activity and discrete selection through gestures.

3. EEG PROCESSING AND MACHINE LEARNING

The EEG processing module is implemented in Python as a multi-threaded bridge connecting OpenBCI, Wekinator, and Ableton Live. The OpenBCI Cyton system provides three data streams: raw EEG signals, acquired at 250 Hz from eight active channels (Fp1, Fp2, C3, C4, P3, P4, O1, O2), spectral band power values, and accelerometer data. Incoming samples are stored in circular buffers of 250 samples (1 s at 250 Hz) that maintain a short temporal window of neural activity. Several signal descriptors are computed from these buffers. In equation (1) the root mean square (RMS) estimates signal energy and is mapped directly to Ableton Live as a continuous control parameter:

$$E = \sqrt{\frac{1}{N}\sum x^2} \tag{1}$$

where N is the number of buffered samples (250) and x are the EEG amplitude values in μV. Frequency bands are defined as: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz). In equation (2) the ratio between theta and alpha power, which following Klimesch [12] is associated with relaxation or meditative states, constitutes the primary spectral feature:

$$\text{relaxation} = \frac{\theta}{\alpha + \epsilon} \tag{2}$$

where ϵ is a small positive constant for numerical stability. In equation (3) the engagement index, adapted from the biocybernetic index of Pope et al. [13] and defined as the ratio between beta activity and the sum of alpha and theta power, provides a complementary descriptor of cognitive arousal:

$$\text{engagement} = \frac{\beta}{\alpha + \theta + \epsilon} \tag{3}$$

These features are transmitted to Wekinator, where two supervised kNN models ($k = 3$) are implemented. The classification model receives two input features: the exponentially smoothed theta/alpha ratio ($\lambda = 0.05$) and a blink-rate estimate derived from delta-band fluctuations. The classifier discriminates exactly two mental states, defined operationally as an eyes-closed concentration state and an eyes-open alert state: in the eyes-closed state the note flow generated by the EEG layer is halted, whereas in the eyes-open state the performer concentrates on the face-tracking interface. The regression model receives six spectral features: slow theta/alpha ratio, relative alpha and theta power, engagement index, signal stability, and an inverted blink-rate term. Classification output is stabilised through a hysteresis filter of 60 consecutive frames before transmission. In equation (4) regression output is further smoothed with a second exponential moving average ($\lambda = 0.02$):

$$\text{EMA}_t = \lambda x_t + (1 - \lambda)\,\text{EMA}_{t-1} \tag{4}$$

where x_t is the current input value at time t. The resulting parameters are mapped to musical controls in Ableton Live through OSC messages.

4. FACE-TRACKING INTERACTION INTERFACE

The second interaction layer is based on facial gesture tracking using MediaPipe FaceMesh [11], which detects 468 facial landmarks from a webcam video stream. The position of the nose tip (landmark index 4 in the MediaPipe topology) is used as a pointer controlling a circular graphical interface composed of several concentric rings (Fig. 2). Each ring corresponds to a different musical parameter, including root note selection, musical scale type, arpeggiator pattern and tempo division. Cursor movement is derived from the normalised coordinates of the nose landmark and smoothed to reduce jitter.

The circular layout was chosen because a radial arrangement keeps every target at a comparable angular distance from the centre, so that all regions are equally and quickly reachable by the nose pointer without large head excursions. The placement of parameters across rings follows their rate of use: tempo divisions, which correspond to the most frequently changed arpeggiator rates, are positioned on the outer ring for the most direct access, while less frequently changed parameters occupy the inner rings.

Interaction relies on a dwell-time mechanism: when the cursor remains inside a region for 0.5 seconds, the corresponding parameter is activated. Although dwell-time selection is sometimes considered unsuitable for real-time musical interaction because of its intrinsic latency [7], in the present system the slow, partly involuntary dynamics of the EEG-driven note flow make this latency musically

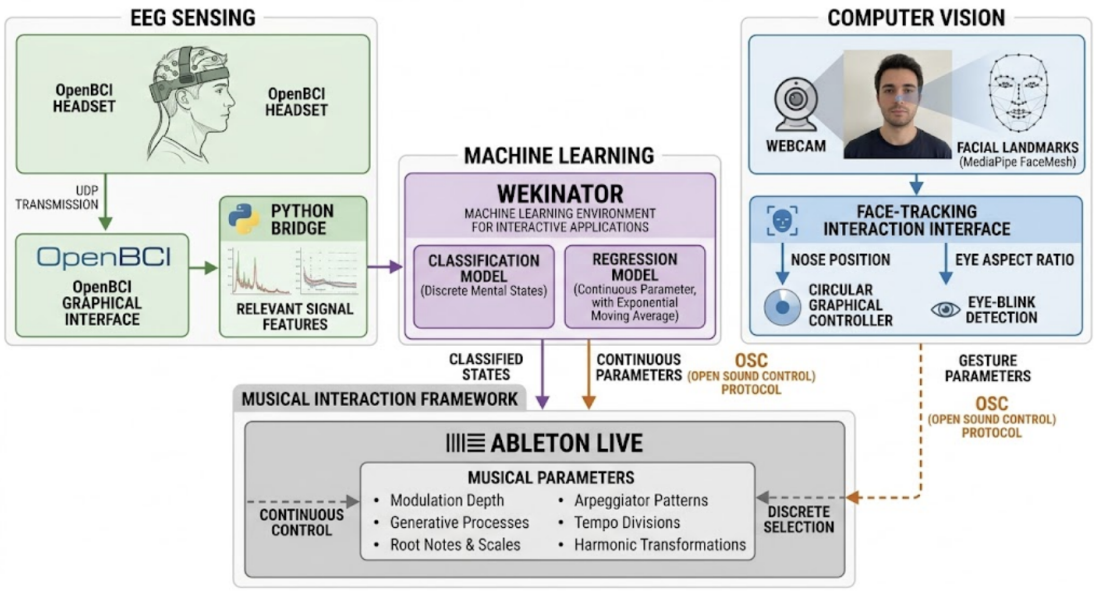


Figure 1. System architecture integrating EEG acquisition, machine learning analysis and facial gesture interaction for real-time musical control.

acceptable: the design prioritises musicality over instantaneous triggering, and the note stream produced by the neural layer remains fast and nearly immediate, so that discrete selections via dwell-time do not disrupt the perceived responsiveness. In equation (5), eye blinks are detected using the Eye Aspect Ratio (EAR) method [14], allowing additional musical actions such as triggering harmonic transformations or confirming parameter selections:

$$\mathrm{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2 \cdot \|p_1 - p_4\|} \tag{5}$$

where p_1, p_4 are the outer eye corners and p_2, p_3, p_5, p_6 are the upper and lower eyelid landmarks. A blink is detected when EAR falls below a threshold of 0.20 for at least two consecutive frames. Beyond blink detection, the gesture vocabulary has been extended with two further FaceMesh-derived gestures: mouth aperture, used to control note density, and a smile gesture, used to activate an additional prepared-piano layer, thereby moving the interface beyond a purely pointer-based mechanic toward a more nuanced gestural expression.

5. PRELIMINARY RESULTS

Preliminary tests were conducted by a single participant using an OpenBCI Cyton headset operating at 250 Hz over eight EEG channels and a webcam-based FaceMesh interface running on a laptop. The system operated in real time with OSC communication between Python, Wekinator and Ableton Live. The total system latency (TSL)—measured from the instant a movement is detected to the activation of the corresponding musical parameters in Ableton Live, and therefore inclusive of webcam capture, landmark inference, feature extraction, OSC transport and audio buffering—remained below 10 ms, confirming the real-time responsiveness of the complete pipeline rather than of the OSC transport alone.

The kNN classifier achieved stable discrimination between the two target mental states (relaxed and focused) across repeated calibration sessions, with training performed on approximately 30 annotated samples per class—a dataset size consistent with the preliminary scope of the study. The regression model was trained on approximately 3,200 feature vectors extracted by the Python bridge from the OpenBCI stream and transmitted via OSC to Wekinator, producing continuous output values in the range $[0, 1]$ subsequently mapped to Ableton Live parameters.

Regarding the use of Wekinator outputs, the two classified states are mapped to a discrete gate that enables or halts the EEG-driven note flow, whereas the continuous regression output (*depth*) is mapped to modulation depth within generative devices in Ableton Live; the specific mappings were selected empirically during calibration so as to keep the most perceptually salient sonic changes under continuous neural control. The system was evaluated through free exploration rather than a constrained task, consistently with the preliminary, single-participant scope of this study. Beyond the continuous note flow, musical intent is made observable through the voluntary switching between the two classified states: the eyes-closed concentration state corresponds to a deliberate deactivation of the instrument, in which the data flow is intentionally halted, whereas the eyes-open active-concentration state re-enables the flow. Because this binary transition is under conscious control of the performer, it provides a discrete, intention-bearing layer that can be distinguished from the accidental fluctuations of the continuous EEG signal, thereby separating deliberate musical gestures from involuntary signal variability.

Initial experiments indicate that the hybrid interaction model enables stable and responsive control of musical pa-

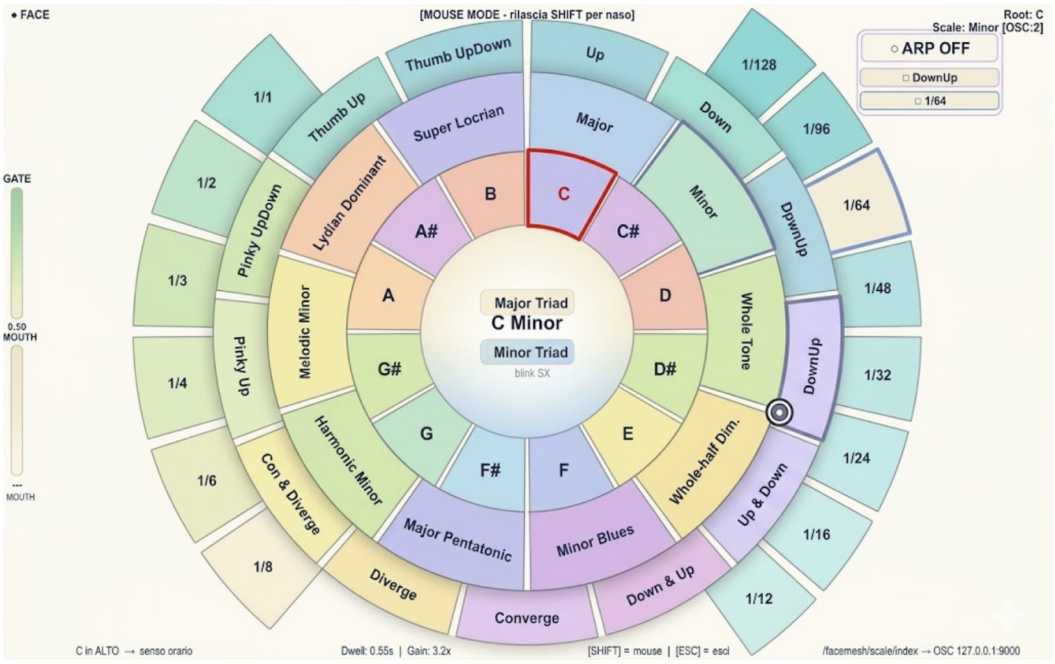


Figure 2. Face-controlled circular interface for selecting musical parameters in real time.

rameters in real time. EEG-derived features provide continuous modulation of musical processes, while the facial interface allows reliable selection of structural elements such as scales or rhythmic patterns. The dwell-time mechanism (0.5 s) proved effective in preventing accidental activations while remaining sufficiently reactive in informal testing, and can be further shortened if a more immediate response is required, and the EAR-based blink detector (threshold 0.20) reliably discriminated voluntary blinks from spontaneous ocular movements in informal testing. In informal trials carried out by the author wearing the headset, the cognitive demand imposed by the EEG layer proved minimal: since the neural control is largely involuntary and is reduced to a binary eyes-closed/eyes-open distinction, it does not require sustained mental effort, so that the performer can concentrate almost entirely on the circular face-tracking interface while the EEG layer operates in the background.

These observations suggest that combining neural signals with facial gestures can increase robustness and usability compared to purely EEG-based musical interaction. Future evaluations will involve formal user studies to assess usability and performance across a broader participant sample.

6. CONCLUSIONS

The proposed architecture demonstrates how hybrid biosignal interfaces can support accessible digital musical instruments. Framed within Waters' concept of performative ecosystems [5], the system positions the user's physiological state as an active agent within an expanded musical network, rather than merely a source of control signals. By combining neural interaction with facial gestures, the system provides both continuous and discrete control mechanisms.

Future work will focus on improving EEG feature extraction, expanding machine learning models, conducting formal user studies including structured subjective evaluations based on questionnaires, think-aloud protocols and task-based assessments with a larger number of users, and integrating additional biosensing modalities—including a respiratory belt and an ECG sensor—to evaluate usability and expressivity across inclusive performance environments. A further direction concerns the generative layer itself: we are currently extending the system with a symbolic generative engine supporting multiple, selectable improvisational styles (e.g. a jazz mode driven by a Markov model trained on the Weimar Jazz Database [15]), so that the performer can switch style in real time through the same dwell-time interface already used for the other musical parameters.

7. REFERENCES

[1] D. Rosenboom, *Biofeedback and the Arts: Results of Early Experiments*. Aesthetic Research Centre of Canada, 1976.

[2] E. R. Miranda and A. Brouse, "Toward direct brain-computer musical interfaces," in *Proc. Int. Conf. New Interfaces for Musical Expression (NIME)*, Vancouver, 2005, pp. 216–219.

[3] T. O. Zander and C. Kothe, "Towards passive brain-computer interfaces: applying brain-computer interface technology to human systems in high workload environments," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025005, 2011.

[4] A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays, "Webgazer: Scalable webcam

eye tracking using user interactions," in *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI)*, New York, 2016, pp. 3839–3845.

[5] S. Waters, "Performance ecosystems: Ecological approaches to musical interaction," in *Electroacoustic Music Studies Network Conference*, 2007.

[6] E. Frid, "Accessible digital musical instruments—a review of musical interfaces in inclusive music practice," *Multimodal Technologies and Interaction*, vol. 3, no. 3, p. 57, 2019.

[7] N. Davanzo and F. Avanzini, "Design concepts for gaze-based digital musical instruments," in *Proc. 19th Sound and Music Computing Conf. (SMC)*, 2022, pp. 472–478.

[8] M. Wright and A. Freed, "Open sound control: A new protocol for communicating with sound synthesizers," in *Proc. Int. Computer Music Conf. (ICMC)*, Thessaloniki, 1997, pp. 101–104.

[9] OpenBCI, "OpenBCI: Open-source biosensing tools," https://openbci.com, 2024.

[10] R. Fiebrink and P. R. Cook, "The Wekinator: a system for real-time, interactive machine learning," in *Proc. Int. Society for Music Information Retrieval Conf. (ISMIR)*, Utrecht, 2010.

[11] C. Lugaresi, J. Tang, H. Nash *et al.*, "MediaPipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[12] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Research Reviews*, vol. 29, no. 2–3, pp. 169–195, 1999.

[13] A. T. Pope, E. H. Bogart, and D. S. Bartolome, "Biocybernetic system evaluates indices of operator engagement in automated task," *Biological Psychology*, vol. 40, no. 1–2, pp. 187–195, 1995.

[14] T. Soukupová and J. Čech, "Real-time eye blink detection using facial landmarks," in *Computer Vision Winter Workshop (CVWW)*, Rimske Toplice, 2016.

[15] M. Pfleiderer, K. Frieler, J. Abeßer, W.-G. Zaddach, and B. Burkhart, "Inside the jazzomat: New perspectives for jazz research," *Schott Campus*, 2017.