Emotion Driven Dialogue — A Model of Mediated Musicianship through Biofeedback-Informed Improvisation

Lluis Guerra Recas*
Independent Researcher, Barcelona, Spain
CESEM / NOVA FCSH
Interactive Technologies Institute (ITI), LARSyS
Lisbon, Portugal
lluisguerra@gmail.com

Olivier Jambois
Department of Music
ENTI - Universitat de Barcelona
Barcelona
olivier.jambois@enti.cat

Abstract

Emotion Driven Dialogue presents an AI-mediated model of musicianship in the form of a real-time improvisational duo setting. The artistic proposal explores how a pianist's embodied signals can become part of a shared musical dialogue with a guitarist, using the Biofeedback Suite (BFS) 2.0 as the mediating system.

BFS 2.0 integrates physiological (GSR), gestural (motion capture), and affective (facial expression) data into real-time sound synthesis and processing. Rather than treating these estimates as definitive emotion classifications, the system reintroduces embodied states into the sonic domain through a mapping framework. The system incorporates an AI-based affective profiling layer designed to derive a personalised valence/arousal baseline and inform initial timbral conditions.

The paper examines how biofeedback-driven mappings influence interaction dynamics, performer responsiveness, and system-mediated musical behaviour. It argues for a non-generative role of AI as an interpretative layer that shapes the conditions of sonic response.

1 INTRODUCTION

Current research on AI-based affective music systems is largely framed around generative approaches that produce musical material, manipulate musical features, or automate aspects of composition in relation to target emotional states. Dash and Agres (2024), for instance, describe AI-based affective music generation systems as systems in which target emotion identification informs an affective music generation component, which then composes or tailors musical output according to discrete emotions or valence–arousal values. However, AI can also operate in less directly generative ways: by mediating the conditions through which musical interaction unfolds. This paper explores such a non-generative role of AI through Emotion Driven Dialogue, a biofeedback-informed improvisation system in which affective profiling shapes the initial sonic conditions of performance without generating musical content directly.

Emotional and gestural processes play a central role in musical performance. Recent advances in sensing and interactive computing enable real-time capture of physiological dynamics (e.g., GSR) and motion, yielding new possibilities for artistic research and technologically mediated performance (Leman, 2007), (Jaimovich and Knapp, 2009). Building on a prior version of an emotion-driven interactive music system for videogames based on music selection (Guerra Recas, 2021a) this project extends the system towards a more robust, configurable platform for biofeedback-mediated composition and improvisation. However, integrating these signals into real-time musical interaction introduces several challenges, including signal noise, hyper-reactivity, system instability, and difficulties in achieving musically coherent behaviour.

In this work physiological and gestural data are approached as embodied signals that can be used in two ways. First, by being directly reintroduced into the musical domain as active sound-generating elements within the interaction. Second, by being processed through an analytic layer that shapes and reconfigures the sonic output based on dynamic affective profiling and historical data analysis. This approach effectively introduces additional musical information, enhancing emotional significance and fostering embodied awareness within the improvisational process.

2 RELATED WORK / BACKGROUND

The use of physiological and gestural data in music performance has a rich history in computer music, human-computer interaction, and embodied cognition. Early studies such as Dainow (1977) documented the physical and motor responses elicited by music. Empirical studies indicate that physiological processes are central to the experience of music, shaping emotional perception and engagement (Baumgartner et al., 2006), (Gabrielsson and Wik, 2003), (Scherer, 2004). Theories of embodied music cognition (Leman, 2007), (Cox, 2016) and emotional body cognition (Juslin and Västfjäll, 2008), (Maes, 2016), provide a conceptual framework for understanding how bodily states, motion, and affect are integral to musical experience and expression.

Early explorations of biofeedback in music include (Tanaka, 2000), who demonstrated the artistic and expressive potential of sensor-based instruments. Recent research highlights the role of emotional and embodied processes in interactive music systems (Volioti et al., 2018), (Wang and Lin, 2025). Jaimovich and Knapp (2009) provided a comprehensive review of physiological signals in performance control, highlighting challenges in signal reliability and mapping strategies. Recent advances in affective mu-

*lluisguerra.com, lluisguerra@gmail.com

sic generation within interactive system contexts (Dash and Agres, 2024) and gestural control (Volioti et al., 2018) have expanded the co-creative possibilities for performer-system interaction, with AI-driven music systems influencing musical processes beyond traditional control or creation paradigms. In previous research, the use of GSR as a direct input to modulate the musical experience proved to be an effective and usable system Guerra Recas (2021b), although emotional data was primarily used for music selection.

The integration of physiological and gestural data into musical performance has long been explored predominantly using such data as control inputs or as sources for emotion detection and analysis (Eaton, 1971), (Bernhardt and Robinson, 2008). In contrast, this work approaches biofeedback as a form of mediated musicianship, where embodied signals are reintroduced into the musical domain as active agents within the improvisational musical discourse. The system emphasizes affective profiling as a way to make emotional processes more visible. Through mapping design, embodied states are rendered sonically within the performance.

BFS 2.0 builds on these foundations, emphasizing performer-specific calibration, adopting multimodal integration as a key to data-experience accuracy (Mauss and Robinson, 2009) - including motion as a key channel for embodied expression (Bernhardt and Robinson, 2008) and real-time adaptability. The motivation of this work focuses on innovation in multimodal interaction and real-time music creation processing within embodied music cognition.

3 SYSTEM OVERVIEW (BFS 2.0)

BFS 2.0 uses biosensors and bodily data within a multimodal setup, including galvanic skin response (GSR), video-based motion tracking, and facial expression descriptors. GSR data is acquired through a low-cost sensor interfaced with an ESP32 microcontroller and transmitted via WebSocket. Motion is captured using a lightweight computer vision approach based on frame differencing, while facial data is extracted using a real-time pipeline based on the DeepFace framework (Serengil and Ozpinar, 2020) (fig. 1).

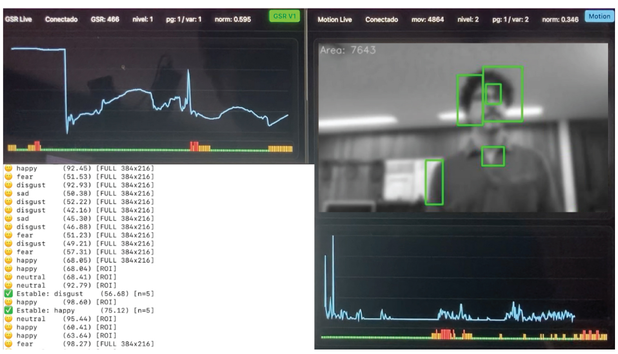

Figure 1: BFS 2.0 monitoring interface, integrating real-time GSR, motion, and facial-expression data. The screenshot shows signal monitoring, activation levels, motion detection, and facial-expression descriptors used during performance.

The system operates through two complementary modes of interaction. On one hand, GSR and motion signals are used as raw material directly mapped to sound parameters, generating continuous sonic output through frequency-based control (fig. 2).

In parallel, incoming signals undergo additional processing. First, a unified stabilization framework—including temporal smoothing, adaptive normalization, and state-transition mechanisms such as hysteresis and refractory periods—ensures perceptual continuity and robustness against noise. This process produces three discrete activation levels (0, 1, and 2), which are mapped to sound effects processing (e.g., delay and reverb), shaping the character of the sound according to the intensity of bodily activity.

At the same time, facial descriptors are used to constrain the harmonic space through scale and chord quantization, shaping the frequencies generated by the reactive layer. Different detected affective states are associated with distinct harmonic frameworks, influencing the melodic behaviour of the system.

Finally, an on-demand hybrid profiling process aggregates multimodal data over a ten-second window to produce an affective-state descriptor and categorical sound profile. The process is deliberately triggered by the pianist rather than operating continuously.

Rather than aiming at precise emotion classification, the system produces symbolic interpretation —including qualitative states and sound profiles— that are then mapped into sound. This process combines heuristic models grounded in psychophysiology combined with AI-based multimodal profiling, enabling both stable baseline estimation and context-sensitive interpretation.

The sound generation layer is implemented in Pure Data, where processed signals and mapping configurations are translated into real-time audio output. Communication between system components is handled via a Python middleware using OSC and MIDI protocols, ensuring flexible and modular integration.

4 CALIBRATION AND PERFORMER PROFILING

BFS 2.0 implements continuous adaptive calibration to account for variability across performers and sessions. Sensor thresholds and activation levels are applied to GSR and motion data and are dynamically computed using recent signal history, including sliding-window normalization and tercile-based segmentation. These values are further stabilized through hysteresis margins and refractory periods, ensuring robust detection of meaningful variations while minimizing noise and rapid state oscillations.

This process results in three discrete activation levels for both GSR and motion signals. In the current implementation, these levels are mapped to sound processing layers applied to the generated audio (level 0 = dry signal, level 1 = stereo delay, level 2 = reverb). This mapping enhances the perceptual intensity of the sound in relation to the performer's physiological arousal, enabling a coherent translation of embodied states into sonic dynamics.

Emotional processes can be observed across multiple temporal scales (Yoon and Kim, 2025). In addition to continuous calibration, BFS 2.0 provides an on-demand affective-profiling mechanism deliberately triggered by the pianist. When the performer presses the "r" key, the system records a ten-second analysis window. Rather than sending the raw

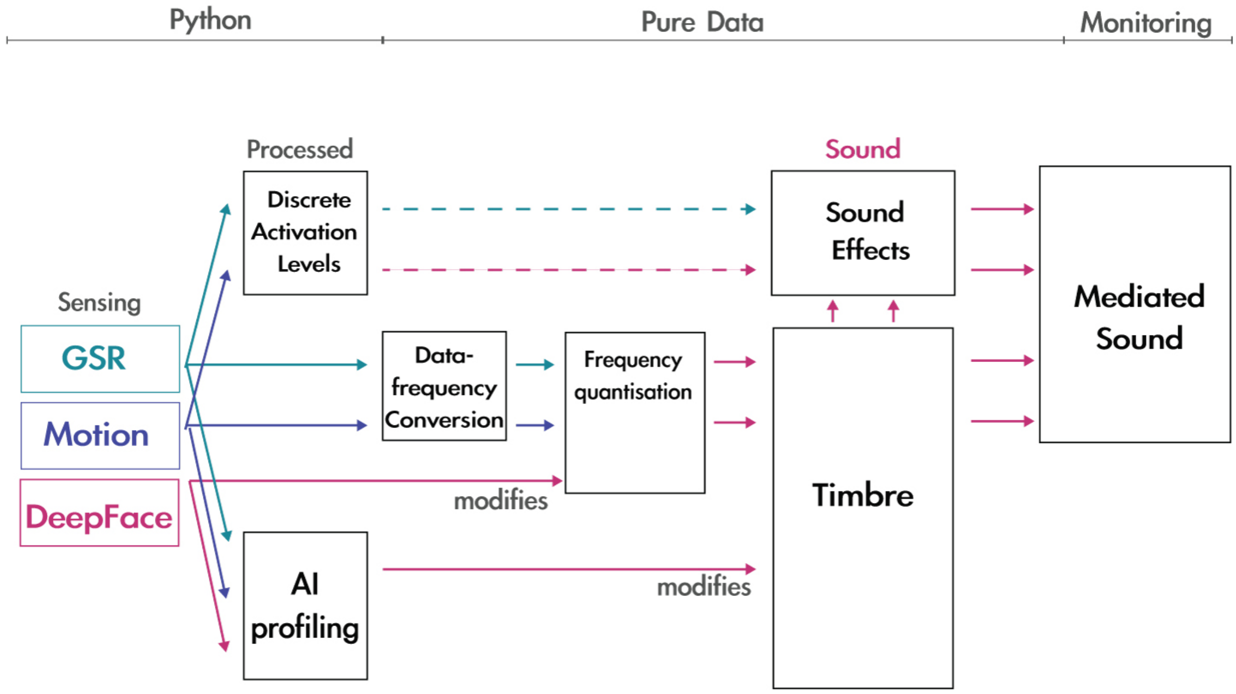

Figure 2: Overview of the BFS 2.0 architecture. Raw GSR, motion, and facial-expression data are acquired in Python and processed through parallel pathways. GSR and motion are mapped to data-frequency conversion and discretised activation levels, which control timbral parameters and sound effects in Pure Data. Facial-expression descriptors extracted with DeepFace contribute to quantisation and to an AI-based affective profiling layer, which produces categorical sound profiles intended to inform timbral and mapping configurations.

time series to the language model, the system locally computes the mean GSR level, normalized GSR, mean motion level, normalized motion, and most frequently detected facial-emotion category.

These descriptors are first converted into a deterministic affective baseline. Arousal is calculated as a weighted combination of normalized GSR and motion (0.7GSR + 0.3Motion), reflecting the stronger contribution assigned to electrodermal activity. Valence is primarily derived from a fixed mapping of the dominant facial-emotion category: happy = 0.8, surprise = 0.7, neutral = 0.5, sad = 0.3, fear = 0.25, and angry or disgust = 0.2. When normalized GSR exceeds 0.7, clearly positive or negative valence estimates are intensified by 0.1. Valence and arousal are then situated within the Circumplex Model of Affect to obtain an initial affective-state descriptor.

The baseline values and summarized sensor descriptors are subsequently included in a structured prompt submitted to GPT-3.5-turbo through the OpenAI Chat Completions API. The model is instructed to preserve the heuristic baseline unless it detects a clear inconsistency and to limit any adjustment of valence or arousal to ±0.2. It returns a JSON object containing refined valence and arousal values, an affective-state label, a confidence estimate, a short poetic description, and a categorical sound profile specifying timbre, rhythm, pitch range, and texture. Successfully parsed responses are transmitted to Pure Data via OSC. If the model is unavailable or its response cannot be parsed, the system uses a deterministic local fallback.

This hybrid profiling layer was explored during the second improvisation session as an on-demand mechanism rather than as a continuously operating controller. The pianist determined when to initiate each profile by triggering the ten-second capture. Although the resulting descriptors and sound profiles were successfully transmitted to Pure Data, their capacity to produce clearly differentiated and musically effective changes was not systematically evaluated. The present study therefore focuses primarily on the integration of the direct biofeedback mappings, while a comparative evaluation of the AI-derived profiles remains for future work.

5 PERFORMANCE DESIGN

The approach of this project is to maintain a performance environment where traditional instruments and improvisational interpretation remain at the core of the musical experience. The setup features a duo format—piano and guitar—where the pianist's bio-signals are brought to the BFS biofeedback system and used as an extension of the piano performance, while the guitarist performs with his own sound palette and a custom Pure Data setup running on a Raspberry Pi. The piano performer also controls the volume levels of the BFS biofeedback system, allowing the pianist to decide whether the "emotional" sounds come into the sound palette or not. Rather than imposing a fixed score, the musical form emerges organically from the interplay between acoustic gestures and the dynamic mappings generated by the BFS system on the piano side, and the independent electronic processing and improvisational choices of the guitarist.

The stage is arranged to support this dialogue: sensors, cameras, and a laptop running BFS and Pure Data are integrated discreetly alongside the instruments and audio interface, ensuring that the focus remains on the musicians' interaction (fig. 3). The central artistic question is how the additional electronic sounds—shaped in real time by the performers' physiological and emotional states—can be meaningfully woven into the ongoing musical conversation. The BFS system is thus conceived not as a dominant layer, but as a responsive participant, enriching the improvisational exchange and opening new expressive possibilities without overshadowing the primacy of instrumental performance.

6 EVALUATION AND OBSERVATIONS

The system was explored through two one-hour duo improvisation sessions involving the same pianist and guitarist. GSR electrodes were attached to the middle and ring fingers of the pianist's left hand, while a camera facing the pianist provided motion and facial-expression data. The first session focused on the direct physiological, gestural, and facial mappings of the biofeedback system. During the second session, the hybrid heuristic–GPT profiling layer was additionally explored through performer-triggered ten-second captures. Several improvisations were recorded in audio. The observations reported below derive from the performers' experience and their retrospective critical listening to these recordings.

6.1 *Supplementary Audio Material*

Two audio excerpts are provided as supplementary material. Recording A documents the main duo improvisation setting and is used as the primary reference for the artistic and interactional observations discussed in this section. In this recording, the biofeedback layer is integrated into the piano–guitar dialogue without systematic timbral reconfiguration, allowing the musical interaction and the stability of the mappings to be heard clearly. Recording B documents a series of exploratory sound experiments using a specific MIDI-based mapping to a Korg Minilogue synthesiser. Although Recording B was produced under less controlled audio conditions, it is included as technical-practice documentation of the system's broader sonic possibilities rather than as the primary artistic outcome.

6.2 *Performer Perspective*

From the pianist's perspective, the integration of biofeedback introduces a novel layer of musical agency. Motion data, in particular, provides an additional expressive channel that operates independently from traditional instrumental technique. This enables the performer to engage in a form of extended embodiment, where bodily gestures directly influence the electronic sound layer. These gestures may be connected to instrumental playing or may operate as standalone bodily movements, suggesting the possibility of extending the system to additional performers or participants.

GSR electrodes were attached to the middle and ring fingers of the pianist's left hand. This placement did not produce perceptually disruptive instability in the mapped output during the two sessions, but it constrained the pianist's instrumental movement. The performer adopted improvisational strategies that minimized displacement and rapid activity in the left hand while continuing to use it musically. This adaptation was feasible for the experienced pianist participating in the study; however, the configuration may be less suitable for techniques requiring extensive or rapid movement in both hands and should be evaluated with additional performers.

Contrary to initial expectations—where small variations or noise in physiological signals are typically assumed to produce perceptually disproportionate sonic responses if not properly stabilised—the combined use of GSR and motion was experienced by the performers as sufficiently stable and musically meaningful during the two sessions. During retrospective critical listening, the performers perceived a relationship between the speed and intensity of movement, the motion-driven sonic output, the perceived musical intensity, and the corresponding GSR-driven sound, supporting the perceptual coherence of the mapping design.

Across the two sessions, the performers reported an increasing ability to anticipate and integrate the system responses, suggesting that BFS 2.0 may function as a learnable instrument rather than a passive interface.

A key observation is the emergence of a recursive feedback loop: when physiological signals are rendered audible, they not only reflect internal states but also influence them. This creates a dynamic cycle in which emotion shapes sound, and sound in turn modulates emotion. This recursive relationship was identified by the performers during retrospective listening, particularly in moments of increased musical intensity.

From a musical leadership perspective, the possibility of deciding when to introduce the biofeedback-generated sound layer provides an additional level of structural control. When used strategically in time, this layer can reinforce or re-activate the musical discourse, functioning not only as a textural addition but also as a form-related or dramaturgical resource.

From the perspective of the informed co-performer (i.e., the guitarist, who does not generate the biofeedback signals), the system is perceived as an extension of the pianist's expressive output. This reinforces the sense of embodied causality and facilitates integration within the improvisational dialogue.

6.3 *System Behaviour*

From a technical perspective, stability mechanisms (temporal smoothing, adaptive normalization, hysteresis, and refractory periods) proved essential for achieving musically usable behaviour. Without these mechanisms, raw physiological and motion signals resulted in erratic and unstable sonic outputs.

One observed limitation concerns the adaptive normalization process. Occasional resets in the normalization window can produce abrupt changes in the mapped output, perceived as discontinuities in the sound. Future developments may benefit from incorporating longer-term historical models or progressive normalization strategies to mitigate these effects.

The discretization of signals into activation levels (0–2) was found to be particularly effective. It reduced sensitivity to noise while maintaining perceptual responsiveness, enabling consistent mapping to sound processing layers (e.g., dry/delay/reverb). This suggests that hybrid approaches combining continuous input with discrete state transitions can provide a balance between expressivity and robustness.

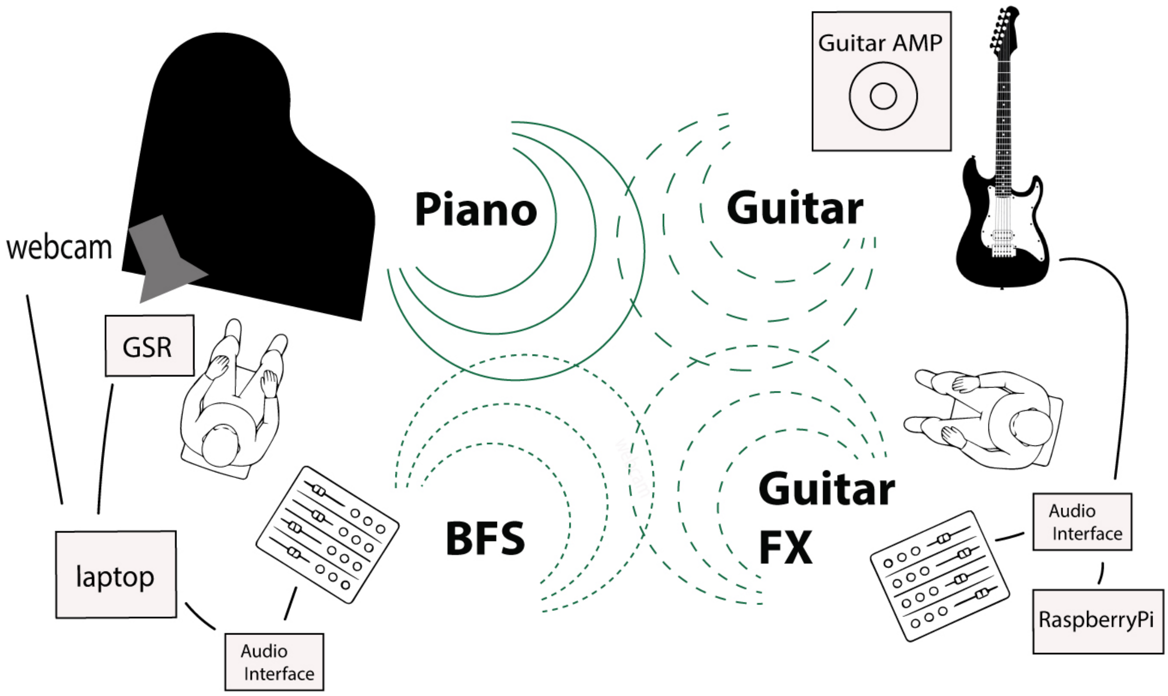

Figure 3: Stage setup for *Emotion Driven Dialogue*. The duo performance combines piano, guitar, and biofeedback-mediated sound. The pianist's biosignals are processed through the Biofeedback Suite (BFS) as an extension of the piano performance, while the guitarist performs with an independent electronic setup based on Pure Data and Raspberry Pi.

Facial expression analysis presents additional challenges in performative contexts. Variations in lighting conditions, performer posture, and the natural constraints of instrumental focus make it difficult to maintain optimal conditions for reliable facial tracking. Furthermore, some performers exhibit relatively neutral facial expressions even during emotionally intense musical states. As a result, the system tends to default toward neutral affective classifications, leading to reduced variability in harmonic mapping.

This highlights the need for improved environmental control (e.g., lighting), more robust tracking methods, or performer adaptation strategies. It also suggests that facial expression alone may not be sufficient as a primary indicator of valence in musical contexts.

At the profiling level, since the snapshot process is performed consciously by the performer, increased attention and intentionality can influence the captured data. This can result in a strong coupling between perceived emotional state and resulting timbral configuration. However, the current mapping between profiles and sound may require further refinement to achieve clearly differentiated and musically meaningful contrasts between distinct affective states.

6.4 *Interaction and Musical Dynamics*

At the interaction level, the BFS system operates as an additional agent within the improvisational ecology. Rather than acting as a deterministic controller, it introduces a semi-autonomous layer that responds to embodied activity while remaining partially unpredictable.

The guitarist reported that the system-generated sounds, when perceived as originating from the performer's embodied state, functioned as a meaningful source of musical inspiration rather than as an external or arbitrary element. This performer report suggests that perceived causality may play an important role in the integration of biofeedback into improvisation.

6.5 *Audience Evaluation as Future Work*

No conclusions regarding audience perception can be drawn from the present study, as no formal audience evaluation was conducted. Moreover, the current configuration provides no visible cues that allow listeners to associate the performer's bodily activity with the resulting electronic sound. Future work should therefore investigate whether audiences perceive the intended signal–sound relationships and whether visual feedback or brief explanatory information improves their understanding of the interaction.

7 DISCUSSION

BFS 2.0 positions itself within the broader context of AI-driven musical systems, particularly in relation to the so-called generative turn, where systems increasingly assume roles of authorship and content production. In contrast, this work proposes a non-generative paradigm in which AI operates as an interpretative and mediating layer rather than as a creative agent.

This contrast can be understood through the comparison between conventional instrumental expression and biofeedback-mediated performance (fig. 4). In the first case, embodied and affective states are expressed primarily through instrumental gesture. In the second, these states are also routed through a mediating system that renders them sonically active within the improvisational ecology.

The difference is not merely technical but conceptual. By avoiding the direct generation of musical material by AI,

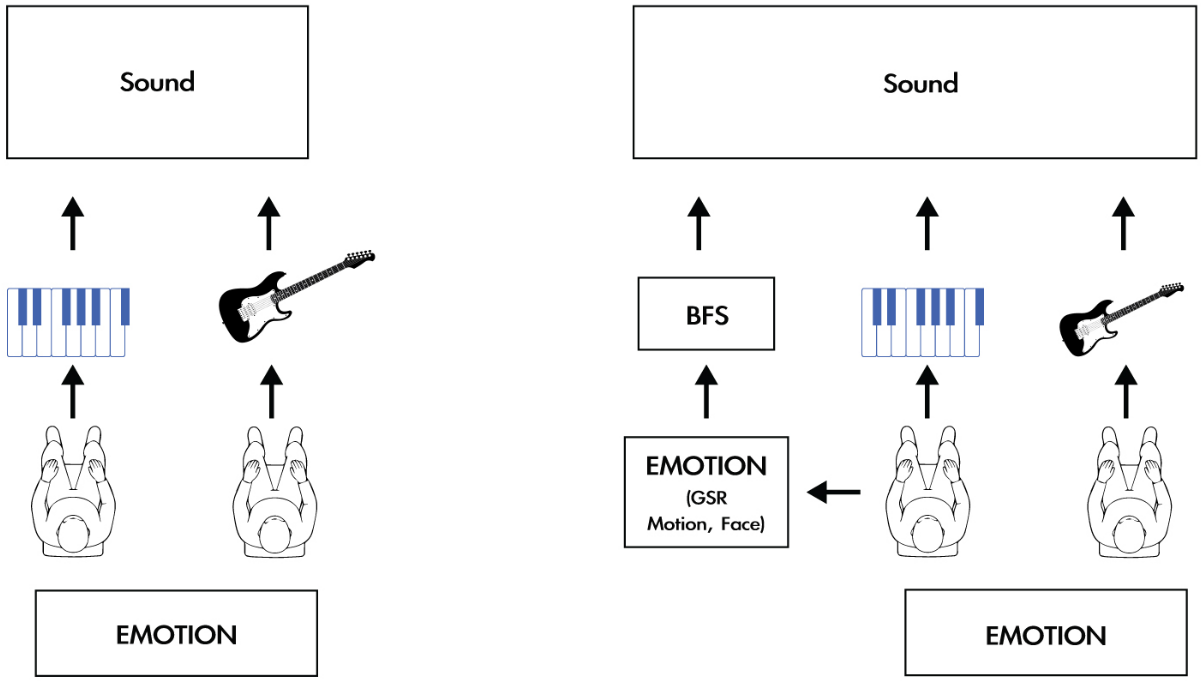

Figure 4: Conceptual model of conventional and biofeedback-mediated performance. In conventional instrumental performance, embodied and affective states are expressed primarily through instrumental gesture. In Emotion Driven Dialogue, the pianist's physiological, gestural, and facial data are additionally routed through BFS, creating a mediated sound layer that participates in the duo interaction without replacing the performers' instrumental agency.

the system preserves the performer as the primary source of musical agency, while AI contributes by shaping the conditions under which embodied signals are rendered sonically. This aligns with a model of mediated musicianship, in which technology extends rather than replaces human expressivity.

The findings suggest that biofeedback-driven systems can function as epistemic tools, making otherwise implicit embodied and affective processes perceptible and actionable within performance. The recursive feedback loop observed—where physiological signals influence sound and are in turn influenced by it—introduces a novel form of interaction that blurs the boundary between internal state and external musical output.

This recursive process may support forms of emotional alignment among performers, particularly when the source of the biofeedback-generated sound is understood by the ensemble. While physiological synchronization was not directly measured in the present study, related phenomena have been documented in collective musical and ritual practices, where shared activity can lead to coordinated physiological and emotional states (Müller and Lindenberger, 2011), (Bernardi et al., 2006), (d'Ausilio, 2015), (Ortiz et al., 2011).

However, several challenges remain. First, the reliability of physiological signals continues to be a limiting factor, requiring careful stabilization and calibration strategies. Second, mapping design is critical: overly direct mappings can lead to trivial or chaotic results, while overly abstract mappings may reduce perceptual clarity and performer control. Achieving a balance between responsiveness, stability, and musical meaning remains an open research problem.

Another key issue concerns transparency. For both performers and audiences, understanding the relationship between embodied signals and sonic outcomes is essential for meaningful engagement. Without this, the system risks being perceived as arbitrary or opaque.

Finally, the present study did not isolate the contribution of the GPT-based refinement through comparison with the deterministic heuristic baseline alone. Future work should therefore include a controlled comparison of both conditions, quantitative analysis of signal–sound relationships, and audience-perception experiments to determine whether the language-model layer provides musically meaningful differentiation beyond the heuristic mapping.

8 CONCLUSION

This paper presented Emotion Driven Dialogue as an AI-mediated model of musicianship in which embodied signals become active elements within a real-time improvisational duo setting. Implemented through BFS 2.0, the system integrates physiological (GSR), gestural (motion), and affective (facial) data to support musically meaningful biofeedback processes.

The practice-based observations suggest that, when supported by appropriate stabilization and mapping strategies, biofeedback signals can be incorporated into live performance in musically meaningful ways. In this context, physiological and gestural data can be effectively treated as expressive dimensions that extend the performer's musical agency.

By repositioning AI as an interpretative layer that shapes the conditions of sonic response, rather than acting as a generator of musical content, AI offers an alternative to prevailing generative paradigms, emphasizing embodiment, interaction, and performer-centered design.

The study also highlights the importance of recursive feedback processes, where sound and physiological state influence each other in real time. This dynamic opens new avenues for understanding musical interaction as an embodied and adaptive system.

Future work will focus on refining mapping strategies, expanding multimodal integration, and developing more rigorous evaluation methodologies. In particular, the inclusion of audience-derived signals and multimodal visualization, including visuals, may further enhance the system's expressive and communicative potential.

Overall, this study contributes to ongoing research at the intersection of embodied cognition, interactive systems, and AI music creativity, proposing a framework in which technology supports—not replaces—the expressive core of musical practice.

9 AUTHOR DECLARATIONS

9.1 *Ethics Statement*

This study involved artistic musical practice using physiological and behavioural signals as interactional material, with no clinical assessment or diagnosis. The automatically generated affective descriptors were used exclusively as artistic interactional material and were not treated as validated measures of the performers' emotional states. Participants gave informed consent for recording and research dissemination.

9.2 *AI Usage Disclosure*

ChatGPT was used for language refinement, structural suggestions, academic phrasing, and assistance with prototype code development and debugging. All research questions, design decisions, analysis, observations, conclusions, and intellectual contributions are the author's own.

REFERENCES

Baumgartner, T., Lutz, K., Schmidt, C. F., and Jäncke, L. (2006). The emotional power of music: How music enhances the feeling of affective pictures. *Brain Research*, 1075(1):151–164.

Bernardi, L., Porta, C., and Sleight, P. (2006). Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. *Heart*, 92(4):445–452.

Bernhardt, D. and Robinson, P. (2008). *Interactive control of music using emotional body expressions*, page 3117–3122.

Cox, A. (2016). *Music and embodied cognition: listening, moving, feeling, and thinking*. Musical meaning and interpretation. Indiana University Press, Bloomington Indianapolis.

Dainow, E. (1977). Physical effects and motor responses to music. 25(3):211–221.

Dash, A. and Agres, K. (2024). Ai-based affective music generation systems: A review of methods and challenges. *ACM Computing Surveys*, 56(11):1–34.

d'Ausilio (2015). What can music tell us about social interaction?

Eaton, M. L. (1971). *Bio-music: biological feedback experimental music systems*. Orcus.

Gabrielsson, A. and Wik, S. L. (2003). Strong experiences related to music: Adescriptive system. *Musicae Scientiae*, 7(2):157–217.

Guerra Recas, L. (2021a). Emotions driven videogame interactive music system. In *International Conference on Human-Computer Interaction*, page 164–169. Springer.

Guerra Recas, L. (2021b). *Personalización subjetiva en los sistemas de música interactiva en los videojuegos. En busca de una experiencia de juego a la medida del individuo*. PhD thesis, Universitat Politècnica de València, Valencia (Spain).

Jaimovich, J. and Knapp, R. B. (2009). Pattern recognition of emotional states during musical performance from physiological signals. *interaction*, 14:15.

Juslin, P. N. and Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31(5):559–575.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*.

Maes, P.-J. (2016). Sensorimotor grounding of musical embodiment and the role of prediction: A review. *Frontiers in Psychology*, 7.

Mauss, I. B. and Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and emotion*, 23(2):209–237.

Müller, V. and Lindenberger, U. (2011). Cardiac and respiratory patterns synchronize between persons during choir singing. *PLoS ONE*, 6(9):e24893.

Ortiz, M. A., Coghlan, N., Jaimovich, J., and Knapp, R. B. (2011). Biosignal-driven art: beyond biofeedback.

Scherer, K. R. (2004). Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them? *Journal of New Music Research*, 33(3):239–251.

Serengil, S. I. and Ozpinar, A. (2020). Lightface: A hybrid deep face recognition framework. In *2020 innovations in intelligent systems and applications conference (ASYU)*, page 1–5. IEEE.

Tanaka, A. (2000). Musical performance practice on sensor-based instruments. *Trends in gestural control of music*, 13(389–405):284.

Volioti, C., Manitsaris, S., Hemery, E., Hadjidimitriou, S., Charisis, V., Hadjileontiadis, L., Katsouli, E., Moutarde, F., and Manitsaris, A. (2018). A natural user interface for gestural expression and emotional elicitation to access the musical intangible cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH)*, 11(2):1–20.

Wang, C.-M. and Lin, C.-H. (2025). Design of an interactive system by combining affective computing technology with music for stress relief. *Electronics*, 14(15):3087.

Yoon, S. and Kim, B. (2025). Multi-scale temporal fusion network for real-time multimodal emotion recognition in iot environments. *Sensors*, 25(16):5066.