Brainwave Entrainment via Generative Audio Models: MusicGen-Based Generation of Alpha-Wave Inducing Music

Oshika Neupane †, Simran Maharjan †, Yugesh K.C †, Atsushi Ito ‡, Yuko Hiramatsu ‡, Anila Kansakar †

† Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Nepal,

‡ Faculty of Economics, Chuo University, Japan

oshika.neupane@example.edu.np

Abstract

We present an EEG-informed generative AI system that fine-tunes MusicGen to produce music associated with increased alpha-band activity (8–13 Hz), a marker of relaxed wakefulness. The pipeline combines consumer-grade EEG recordings, preference-based reward modeling, and LoRA-based REINFORCE fine-tuning. EEG data were collected from 49 participants listening to 43 tracks, with normalized alpha-power changes from baseline used as the reward signal. A pairwise-preference reward model trained on MusicGen hidden-state embeddings reduced loss from 0.90 to 0.34 over 100 epochs, while the RL reward curve increased over 7,000 steps. In the validation set, the fine-tuned model produced a smaller decrease in alpha than the baseline, although the difference was not statistically significant (Welch $t = 0.32$, $p = 0.381$). We also developed a mobile application for portable EEG data collection. These results suggest the feasibility of guiding generative audio models with physiological signals.

Keywords: EEG, Brainwave Entrainment, Generative AI, MusicGen, Reinforcement Learning, Reward Modeling, Alpha Waves, Neurofeedback

1. Introduction

Alpha waves (8–13 Hz) are dominant during relaxed wakefulness and are reliably suppressed by anxiety and cognitive overload (Jiang et al., 2024). Practiced meditators consistently show elevated frontal and parietal alpha power (Takabatake et al., 2021), making alpha elevation a useful proxy for the calm state underlying stress relief and mindfulness practice; we focus on alpha specifically, rather than other EEG bands, because it is the most well-established biomarker of relaxed wakefulness and has the most direct precedent in neurofeedback literature. Transformer-based generative models such as MusicGen (Copet et al., 2024), MusicLM (Agostinelli et al., 2023), and Jukebox (Dhariwal et al., 2020) synthesize high-fidelity music, but optimize purely for stylistic coherence and carry no awareness of neurological impact. No mechanism currently exists within these models to preferentially generate audio that promotes a target brainwave state.

This work pursues two goals: (1) build an EEG data-collection pipeline using the MINDSALL MA900 sensor that derives a scalar reward score from the normalized increase in alpha-band power between baseline and music-listening sessions, and (2) fine-tune a pretrained MusicGen model toward alpha-promoting generation using this reward signal via a pairwise-preference reward model and a LoRA-adapted REINFORCE loop. The proposed system was evaluated using EEG recordings collected from 49 participants across 43 music tracks.

Large-scale generative music models such as MusicGen are trained purely on musical data and carry no awareness of their neurological impact. Conventional brainwave-entrainment approaches instead rely on static, pre-recorded audio that is neither optimized for nor responsive to individual neural responses. Although consumer-grade EEG devices can capture meaningful brainwave data, this signal is rarely leveraged to evaluate or improve generated audio, and the gap between EEG measurement and generative audio modeling remains largely unaddressed.

2. Background and Related Work

EEG band power is a well-validated indicator of affective state (Jiang et al., 2024). The alpha band (8–13 Hz), generated mainly in thalamocortical circuits, is prominent over occipital and parietal regions during eyes-closed rest and increases with relaxed, inward attention. Alpha oscillations reflect cortical idling: as the brain disengages from external demands, sensory input is suppressed, alpha power rises alongside lowered heart rate and muscle tension, and subjective calmness emerges. Practiced meditators consistently show elevated alpha power during and immediately after practice, and alpha enhancement is widely used as an objective neural correlate of the early stages of meditative absorption, even in non-meditators (Takabatake et al., 2021); this positions alpha elevation as a proxy relevant not only to general relaxation but to mindfulness-oriented mental states more specifically. Koelsch et al. showed that pleasant and unpleasant music elicit distinct alpha and theta responses (Sammler et al., 2007), and music-based neurofeedback using real-time auditory feedback has been shown to re-

inforce alpha activity directly (Takabatake et al., 2021).

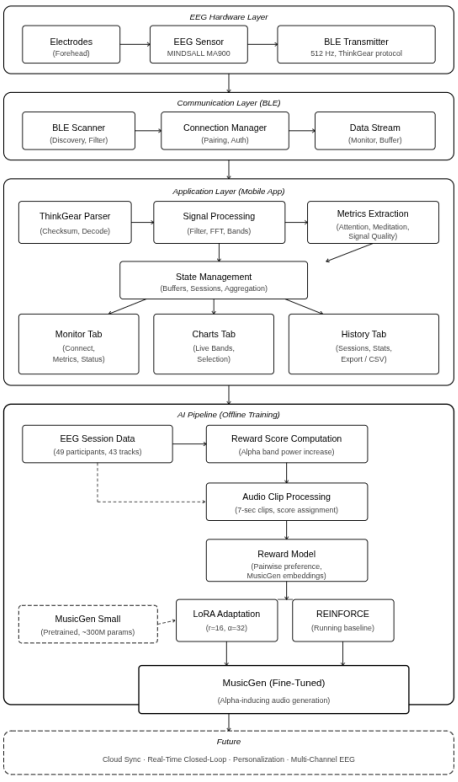
MusicGen encodes raw audio via EnCodec into discrete token sequences, then models them autoregressively with a transformer conditioned on text descriptions; its 300M-parameter "small" variant is computationally tractable for RL fine-tuning, which is why we adopt it as the base model here rather than larger text-to-music systems.

MusicRL demonstrated that RLHF can align music generation with human preference ratings (Cideron et al., 2024). ReaLJam and RL-Duet explored real-time interactive music generation with reinforcement learning (Scarlatos et al., 2025; Jiang et al., 2020), and Dutta et al. used EEG signals to guide RL-based music selection for emotion regulation (Dutta et al., 2020). Our work differs from all of these: we use EEG *physiological* signals, not subjective ratings or discrete emotion labels, to fine-tune the *generative model's parameters* directly, rather than selecting among pre-existing tracks.

3. Methodology

Figure 1 shows the end-to-end pipeline: EEG hardware and communication, a companion mobile application, and an offline AI training pipeline.

Figure 1 End-to-end system architecture: EEG hardware, BLE communication, mobile application, and offline AI training pipeline.



3.1 EEG Data Collection

Sessions used the MINDSALL MA900, a single-channel consumer-grade EEG headband (frontal lobe, 512 Hz, BLE). A Python pipeline decoded ThinkGear packets, applied a second-order 50 Hz IIR notch filter, and computed FFT-based power spectral density (PSD) at 6–14 Hz via a Hamming-windowed 1,000-sample sliding buffer. Each session comprised a baseline phase (eyes closed, no music) followed by a music-listening phase. Data were collected from **49 participants** across **43 music tracks** in diverse real-world environments (classrooms, offices, homes).

To remove the dependency on a laptop during sessions, we additionally built a React Native Android application that mirrors the same decoding and signal-processing pipeline–ThinkGear decoding, notch filtering, Hamming windowing, and FFT-based PSD–and provides participants with real-time brainwave visualization and structured CSV export, validated against the Python pipeline's output by comparing exported band-power and PSD features across matched sessions. The device's built-in meditation index, which aggregates inward attentional focus, additionally served as an auxiliary sanity check against the independently computed alpha-power reward scores.

3.2 Reward Score Computation

The session-level reward score is the normalized increase in mean alpha-band power ($\bar{\alpha}$, averaging Alpha Low and Alpha High) between the music and baseline phases:

$$r = \frac{\bar{\alpha}_{\text{music}} - \bar{\alpha}_{\text{base}}}{\bar{\alpha}_{\text{base}} + \varepsilon}, \quad \varepsilon = 10^{-6}$$

A positive r indicates that the track elevated alpha power relative to the resting baseline, signalling a shift toward the relaxed, internally focused neural condition associated with reduced stress and meditative readiness. One score is computed per session CSV and saved to a consolidated reward dataset used throughout training.

3.3 Audio Clip Processing

Each track was segmented into non-overlapping 7-second clips (≈20–25 per track), chosen as a compromise between preserving enough temporal context for the reward model and keeping token sequences short enough for efficient training. Each clip inherits its session's reward score, then is resampled to 32 kHz mono via Audiocraft's `convert_audio` function, ensuring a consistent format before tokenization by MusicGen's EnCodec compressor.

3.4 Reward Model

Token sequences are truncated to T=256 tokens and passed through MusicGen's frozen language model; the final hidden state $\mathbf{e}_i \in \mathbb{R}^{1024}$, obtained by modifying MusicGen's `lm.py` to return hidden states alongside logits, serves as a compact acoustic representation of each audio clip. A lightweight MLP maps this embedding to a scalar predicted reward:

$$\hat{r} = \mathbf{W}_2 \cdot \text{Dropout}(\text{ReLU}(\mathbf{W}_1 \mathbf{e}))$$

with $\mathbf{W}_1 \in \mathbb{R}^{1024\times32}$, $\mathbf{W}_2 \in \mathbb{R}^{32\times1}$, dropout= 0.3, trained with a margin-based Bradley-Terry pairwise loss:

$$\mathcal{L} = -\log \sigma(\hat{r}_{\text{winner}} - \hat{r}_{\text{loser}})$$

using AdamW (lr=10^{-4}, wd=10^{-2}) with a StepLR schedule (γ=0.9, step= 500) over 100 epochs and an effective batch size of 16. Only the reward-model parameters are updated; MusicGen's language model remains frozen throughout.

3.5 Reinforcement Learning Fine-Tuning

The trained reward model predicts the alpha-inducing potential of newly generated clips, and these predicted rewards serve as the reinforcement learning signal for fine-tuning MusicGen itself. LoRA adapters (r=16, α=32, ≈0.1% of parameters) are injected into MusicGen's attention output projections (`out_proj`) and feedforward layers (`linear1`, `linear2`), while all base weights remain frozen. The REINFORCE update maximizes

$$\mathcal{L}_{\text{PG}} = -\mathbb{E}\big[A_t \cdot \log \pi_\theta(a_t|s_t) + \lambda \mathcal{H}(\pi_\theta)\big]$$

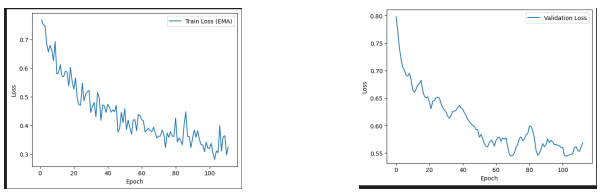
where $A_t = r - b_t$ is the advantage over an exponentially smoothed baseline (α_b=0.6) and λ=0.005 is the entropy regularization coefficient. Sequences of 256 tokens (≈5 s) are sampled at temperature 0.8, top-k 128, with gradient accumulation over 2 steps for 500 iterations, saving checkpoints every 30 iterations. REINFORCE was selected over more complex alternatives such as PPO for its simplicity and suitability given the available computational resources.

4. Results and Discussion

4.1 Data Collection and Reward Modeling

Both the Python pipeline and mobile application maintained stable BLE connections across all 49 participant sessions. Checksum validation silently discarded corrupted packets, and the notch filter effectively suppressed mains interference. CSV export–including all eight band powers, attention, meditation index, signal quality, and nine PSD values at 6–14 Hz–was validated for structural consistency between both platforms.

Figure 2 Reward model training loss (left) and validation loss (right) over 100 epochs.



As shown in Figure 2, training loss fell from 0.90 to 0.34 (a 62% reduction) and validation loss fell from 0.80 to 0.56, indicating that MusicGen embeddings carry sufficient acoustic information for preference ranking. A growing train–validation gap suggests mild overfitting, attributable to the limited number of distinct reward values (43), since all clips from the same track inherit one session-level score.

4.2 RL Fine-Tuning

Figure 3 RL training reward curve over 7,000 steps.

Figure 3 shows an overall upward trend, indicating that the model progressively learns to generate higher-quality outputs: in the early stages (0–1,000 steps) rewards are highly variable, reflecting stochastic sampling and the initially untrained LoRA parameters, before rising steadily to approximately 8–9 by step 3,000, demonstrating effective policy improvement. Two transient drops near steps 1k and 4.2k, likely from sampling-parameter decay, recover quickly, consistent with stable policy-gradient learning under EEG-derived rewards.

4.3 Validation Study

To test whether the fine-tuned model produces more alpha-inducing audio than the untrained baseline, participants listened to music generated by both the pretrained MusicGen model and the fine-tuned model, and the resulting change in alpha power relative to baseline was compared between the two conditions. The mean change in Alpha High power was −5631.67 for the trained model versus −7362.78 for the untrained model, and Alpha Low showed a similar pattern (−3950.47 vs. −4058.74): both trends favor the trained model. A one-tailed Welch's t-test yielded $t = 0.32$, $p = 0.381$, above the $\alpha = 0.05$ threshold, so we cannot reject the null hypothesis. Given the small validation sample, this limits statistical power to detect a modest effect; the direction of the trend is nonetheless consistent with the intended optimization target, and a larger within-subjects study is needed to establish significance. Table 1 summarizes these results alongside the reward-model and RL training outcomes.

4.4 Limitations

Although the observed trends are encouraging, several factors limit the strength of the conclusions. The current system is offline, with rewards pre-computed rather than streamed live; the dataset is comparatively small (49 participants, constrained by data-collection time and resources) and drawn from a single-channel frontal EEG

Table 1 Summary of key results.

Metric	Value
Reward model train loss (start→end)	0.90 → 0.34
Reward model val. loss (start→end)	0.80 → 0.56
RL reward (start→end, 7,000 steps)	−3 → 9
Alpha High Δ (trained vs. untrained)	−5631.67 vs. −7362.78
Alpha Low Δ (trained vs. untrained)	−3950.47 vs. −4058.74
Welch's t-test	$t = 0.32$, $p = 0.381$

device; and the reward signal captures only alpha-band power. Clip-level score inheritance and the small number of distinct reward values are the most direct causes of the overfitting observed above. Addressing these constraints, extending the pipeline toward real-time closed-loop operation and multi-band reward signals, and completing a larger-sample validation study, are the natural next steps for this line of work.

4.5 Cognitive Science Significance

Beyond the engineering pipeline, this work bears on how physiological signals can inform generative modeling of cognitive and affective states. Alpha-band power is a well-established correlate of relaxed wakefulness and, more specifically, of the early stages of meditative absorption: practiced meditators consistently show elevated frontal and parietal alpha during and immediately after practice (Takabatake et al., 2021). Using alpha as a reward signal therefore grounds the optimization target in a physiologically interpretable marker, rather than in subjective self-reports or discrete emotion labels, and potentially reduces the annotation biases associated with human-preference labeling.

The validation study in § 4.3 provides a preliminary, though not statistically significant, link between the optimization target and measured EEG outcomes. The trained model produced a smaller alpha decline than the untrained baseline in both the Alpha High and Alpha Low sub-bands, matching the intended direction of the reward signal. This consistency suggests that generative audio models can, in principle, be guided by explicit physiological objectives rather than indirect proxies such as genre or tempo labels, which we see as the central cognitive-science implication of this work.

We view the present pipeline as a working proof of concept rather than a validated personalized system. Extending it toward user-level modeling, where the reward model and RL fine-tuning adapt to an individual's own alpha response, is a natural next step and would connect this line of work more directly to personalized neurofeedback and mindfulness-oriented applications, beyond general relaxation.

5. Conclusion

We presented an end-to-end pipeline that uses EEG-derived alpha-band power as a reward signal to fine-tune MusicGen through a pairwise-preference reward model and LoRA-adapted REINFORCE. The system combines structured physiological data from 49 participants listening to 43 tracks with a Python-based analysis pipeline and a companion mobile application, and steers generative music toward higher predicted alpha responses. An initial validation study showed a trend in the intended direction, though not statistically significant, motivating larger-scale follow-up experiments. Planned extensions, including cloud synchronization, real-time closed-loop feedback, per-user personalization, and multi-channel EEG, address the limitations identified in Section 4.4 and are summarized in Figure 1. Overall, this work provides a proof of concept for brain-state-aware generative audio modeling, connecting physiological measurement with large-scale generative music systems for relaxation and mindfulness-oriented applications.

Acknowledgement

This research is supported by JSPS Kakenhi(23H03649).

References

Agostinelli, A., Denk, T.I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., Huang, Q., Jansen, A., Roberts, A., Tagliasacchi, M., Sharifi, M., Zeghidour, N., & Frank, C. (2023). MusicLM: Generating Music From Text..

Cideron, G., Girgin, S., Verzetti, M., Vincent, D., Kastelic, M., Borsos, Z., McWilliams, B., Ungureanu, V., Bachem, O., Pietquin, O., Geist, M., Hussenot, L., Zeghidour, N., & Agostinelli, A. (2024). MusicRL: Aligning Music Generation to Human Preferences..

Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., & Défossez, A. (2024). Simple and Controllable Music Generation..

Dhariwal, P., Jun, H., Payne, C., Kim, J. W., Radford, A., & Sutskever, I. (2020). Jukebox: A Generative Model for Music..

Dutta, E., Bothra, A., Chaspari, T., Ioerger, T., & Mortazavi, B. (2020). Reinforcement Learning using EEG signals for Therapeutic Use of Music in Emotion Management. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Vol. 2020, pp. 5553–5556.

Jiang, H., Chen, Y., Wu, D., & Yan, J. (2024). EEG-driven automatic generation of emotive music based on transformer. *Frontiers in Neurorobotics*, *18*.

Jiang, N., Jin, S., Duan, Z., & Zhang, C. (2020). RL-Duet: Online Music Accompaniment Generation Using Deep Reinforcement Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*, 710–718.

Sammler, D., Grigutsch, M., & Koelsch, S. (2007). Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*, *44*, 293–304.

Scarlatos, A., Wu, Y., Simon, I., Roberts, A., Cooijmans, T., Jaques, N., Tarakajian, C., & Huang, C.-Z. A. (2025). ReaLJam: Real-Time Human-AI Music Jamming with Reinforcement Learning-Tuned Transformers..

Takabatake, K., Kunii, N., Nakatomi, H., Shimada, S., Yanai, K., Takasago, M., & Saito, N. (2021). Musical Auditory Alpha Wave Neurofeedback: Validation and Cognitive Perspectives. *Applied Psychophysiology and Biofeedback*, *46* (4), 323–334.