オーガナイズドセッション

2026年8月31日(月) 14:30 ~ 16:30 | 2階 OS会場6(BNC211)

創造性あるいは創造的活動と脳のデフォルトモードネットワークについて考える

オーガナイザー:伊藤 篤（中央大学)、上田 一貴（東京大学)、原田 康也（早稲田大学)、平松 裕子（中央大学)

生成系AIが広く使われるようになってきており、人とAIの関わりについて、関心が高まっている。その中で、人がアイデアを出して、AIに実行させる、という役割分担の可能性がしばしば指摘されている。ここで、アイデアを出すというのは、どういうことだろうか？アイデアを出せと言われても、なかなかすぐには思いつかないのが現状である。近年、マインドワンダリングと脳のデフォルトモードネットワークと創造性の関係に関連する研究が進んでおり、興味を持つ人も増えているが、その実態は、あまり明らかではない。そこで、本OSでは、マインドワンダリングと脳のデフォルトモードネットワークについての理解を深めるため、関連する発表やワークショップを行う。

ワークショップ（PDF）
企画講演1（PDF）
企画講演2（PDF）
企画講演3（PDF）

思考・知識

[OS1-6-01-A] 散歩中のマインドワンダリングはアイデアの創造性を促進するか？

*西尾 愛梨亞[1]、石黒 千晶[2]、横澤 一彦[3] (1. 北海道大学大学院教育学院、2. 東京大学、3. 日本国際学園大学)

感情

[OS1-6-02] Brainwave Entrainment via Generative Audio Models

Neupane Oshika[2]、Maharjan Simran[2]、K.C Yugesh[2]、伊藤 篤[1]、平松 裕子[1]、*Khansakar Anila[2] (1. 中央大学、2. Tribhuvan University)

散歩中のマインドワンダリングはアイデアの創造性を促進するか？

Does Mind-Wandering During Walking Enhance Creative Ideation?

西尾 愛梨亞[†]，石黒 千晶[‡]，横澤 一彦*
Aria Nishio, Chiaki Ishiguro, Kazuhiko Yokosawa

[†]北海道大学大学院教育学院，[‡]東京大学，*日本国際学園大学
[†]Graduate School of Education, Hokkaido University, [‡]The University of Tokyo, *Japan International University
nishio.aria.q7@elms.hokudai.ac.jp

概要

目の前の課題から注意が逸れて，関係のない思考をする現象は，マインドワンダリング（Mind Wandering）として知られる。本研究では，創造性研究で孵化効果を生むと言われる散歩に着目し，意図的にマインドワンダリングを行うことの効用を示すことで，散歩中にマインドワンダリングが促され，散歩後の創造性課題のパフォーマンスが高まるかどうかを検討した。その結果，実験群と統制群の両群とも流暢性が高まったこと，統制群において柔軟性が高まったことが示された。

キーワード：マインドワンダリング(Mind-Wandering)，創造性，散歩，孵化効果，気分

1. はじめに

1.1 背景

誰しも，散歩中に思考が次々と連想されるという経験があるだろう。このように，目の前のことから注意が逸れて，関係のない思考をする現象のことを，マインドワンダリング（Mind Wandering：以下 MW と表記）という（Smallwood & Schooler, 2015）。MW が起こる背景として，脳のデフォルトモードネットワーク（Default Mode Network）と関連することが指摘されている（Raichle, 2015）。創造性研究では，課題から一旦離れて時間を空けると創造性が促進することは孵化効果として知られている。Thabane et al (2026) は，散歩が拡散的思考を増加し，創造的なアイデアの創出を刺激する介入策となる可能性を示唆している。また，MW が創造性に与える効果に関しては，山岡・湯川（2016）らをはじめとした多くの研究がなされてきたが，いずれも一貫した結果は示されていない。

1.2 問いと仮説

本研究では日常生活場面で孵化効果が生まれやすい行動の 1 つである散歩に着目し，散歩中に生じる MW が散歩前後の気分や創造性にどのような影響を及ぼすのかを実験的手法によって明らかにすることを目的とした。参加者を孵化効果に関する教示を行う実験群と教示を行わず自由な MW を統制する統制群に群分けし，孵化効果の教示を受けた実験群の方が，意図的に MW を行うことにより，創造性課題の得点と創造的自己効力感が高まるだろうという仮説を立てた。

2. 方法

2.1 実験計画

本実験は散歩前の介入の有無の要因（散歩前に孵化効果の教示をする実験群と，教示をせずに散歩中にクイズに回答する統制群の 2 水準）および時期の要因（事前事後の 2 水準）の 2 要因混合計画で実施した。従属変数は気分，創造性課題の評価，創造的自己効力感であった。本研究は東京大学倫理審査委員会の承認を得て行われた（承認番号：E2025ALS030）。

2.2 実験参加者

聖心女子大学に所属する女子大学生・大学院生 50 名が参加した。Big Five の開放性が同程度となるように実験群と統制群にそれぞれ 25 名ずつ割り当てた。

2.3 手続き

実験は聖心女子大学構内で 1 人ずつ実施され，所要時間は 1 時間程度であった。散歩前に，実験群には孵化効果に関して「**研究では，散歩などで気分転換をすることで思わぬひらめきが得られることがあると報告されています**」と教示を行い，統制群には散歩中に「**正門前にある，構内走行速度の標識は 8 キロである**」「**宮代会館前のガウンを着たクマは 2 匹である**」などといった 7 問の正誤課題への回答を求め，両群とも 30 分程度の散歩をしてもらった。散歩後に MW のインタビューを行い，散歩前後で質問紙や代替用途課題（Alternative Uses Task：以下 AUT と記載）の回答を求めた（図 1）。

2.3.1 創造性課題

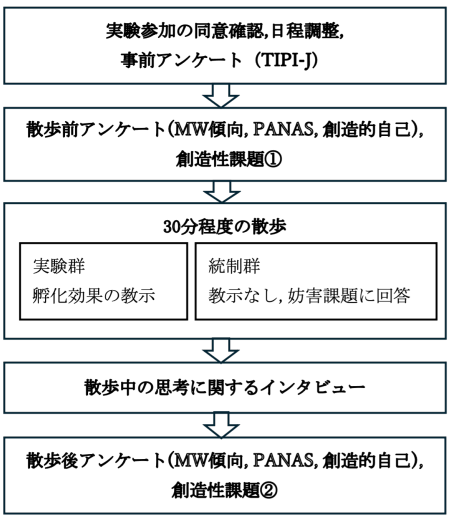
創造性課題として，ものの通常とは異なる使用方法のアイデアを回答する代替用途課題（Alternative Uses Task：以下 AUT と記載）への回答を求めた。参加者には「缶詰の空き缶」「靴下」の刺激を提示し，通常とは異なる使い方を3分間回答するよう求め，Google Form

上でアイデアを箇条書きで記入するよう求めた。AUTの刺激は事前事後の順序効果を相殺するためにカウンターバランスをとった。分析は，先行研究（山岡・湯川, 2016；石黒他, 2025）を参考に，流暢性，柔軟性，独自性の観点から採点された。流暢性は，アイデアの回答数を流暢性得点とした。柔軟性はアイデアをカテゴリー別に分類し，カテゴリー数を柔軟性得点とした。例えば空き缶について，「缶蹴り」と「積み木」は同じカテゴリーとして分類されたが，「缶蹴り」と「植木鉢」は異なるカテゴリーとして分類された。評定者間のカテゴリー数の一致率として相関係数を算出したところ，缶：$r = 0.98$ ($p < .001$)；靴下：$r = 1.00$ ($p < .001$) であった。独自性は人間採点とAI採点を用いた。人間採点は「1: まったく創造的ではない」から「5: 非常に創造的である」の5段階評定を行い，評定者間の一致率は缶：$r = 0.93$ ($p < .001$)；靴下：$r = 0.86$ ($p < .001$) であった。AI採点では，創造性課題の自動採点に関するWebサイトCAP（https://cap.ist.psu.edu）内の，各単語との間の最大意味距離を計算する指標であるSemDisと単語からの応答を評価するために微調整されたニュートラルネットワークであるCLAUSを用いて採点を行い，アイデアごとの回答を，翻訳サイトのDeepLを用いて英訳した（https://www.deepl.com/ja/translator）。

2.3.2 質問紙

質問紙は，日本語版意図的/非意図的マインドワンダリング傾向尺度（山岡・湯川, 2019），日本語版Ten Item Personality Inventory（小塩・阿部, 2012），日本語版Positive and Negative Affect Schedule（佐藤・安田, 2001），日本語版創造的自己尺度（Ishiguro et al., 2021）を使用した。

図1　本実験の手続き



3.　結果

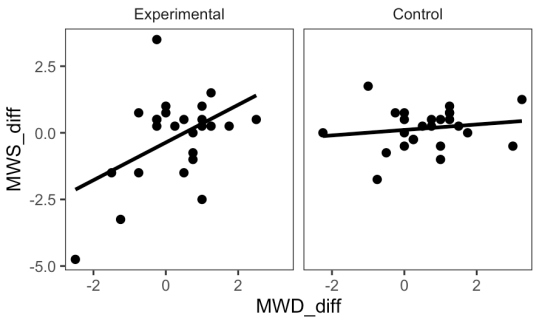
まず，群間で意図的・非意図的マインドワンダリング傾向が異なったかどうかを検討した。次に，創造性関連変数や気分に関する変数について，群（実験群/統制群）と時期（事前/事後）を固定効果，参加者をランダム効果とする線形混合モデル分析をおこなった。

3.1 意図的/非意図的マインドワンダリング傾向

意図的MWでは時期の固定効果が有意であった（$F(1, 48) = 7.36$, $p < .01$）。しかし，群×時期の固定効果の交互作用は有意ではなかった。非意図的MWではいずれも有意ではなかった。両群とも散歩前より散歩後で意図的なMW傾向が上昇したが，非意図的MW傾向には有意な差は見られなかった。

また，両群の意図的MW（MWD_diff）と非意図的MW（MWS_diff）の変化量を示した図が図2である。両群とも意図的MWと非意図的MWの変化は正の関係にあるが（実験群：$r = 0.46$, $p < .05$；統制群：$r = 0.16$, $p < .05$），実験群の方が傾きが大きく，個人差が見られた。

図2　両群の意図的MW（MWD_diff）と非意図的MW（MWS_diff）の関係



3.2 日本語版 Positive and Negative Affect Schedule

ポジティブ気分では時期の固定効果が有意であった（$F(1,48) = 9.21$, $p < .01$）が，群×時期の固定効果の交互作用は有意ではなかった。ネガティブ気分も同様に時期の固定効果が有意であった（$F(1,48) = 7.35$, $p < .01$）が，群×時期の固定効果の交互作用は有意ではなかった。このことから，両群とも散歩後にポジティブ気分が上昇し，ネガティブ気分が低下したことが示された。

3.3 創造的自己効力感

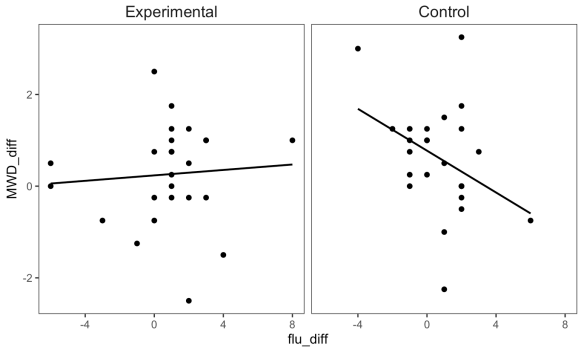
創造的自己効力感はいずれの固定効果も有意ではなかった。

3.4 創造性課題（AUT）

3.4.1 流暢性得点

流暢性得点は群の固定効果が有意傾向であり（$F(1, 48) = 3.25, p < .10$），時期の固定効果も有意であった（$F(1, 48) = 5.06, p < .05$）が，固定効果の交互作用は有意ではなかった（$F(1, 48) = 0.03, p = .86$）。このことから，散歩後にアイデア回答数が全体的に増加したものの，介入の独自的な効果は示されなかった。また，両群の意図的 MW と流暢性得点の変化量を示した図 3 からは，統制群において流暢性得点が上がるほど意図的 MW が下がっていたことが分かる。

図 3　両群の意図的 MW（MWD_diff）と流暢性得点（flu_diff）の関係



3.4.2 柔軟性得点

柔軟性得点は時期の固定効果が有意であり（$F(1, 48) = 5.26, p < .05$），群×時期の固定効果の交互作用が有意傾向であった（$F(1, 48) = 2.87, p < .10$）。探索的に単純主効果を検討した結果，実験群よりも統制群の変化量が大きく（実験群：0.12；統制群：0.80），統制群においてのみ散歩前と散歩後に有意な差が認められた（$t(48) = 2.82, p < .01$）。従って，統制群において散歩後に柔軟性得点が高くなったことが示された。

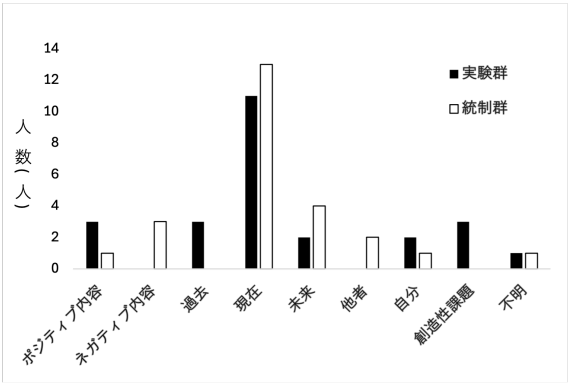
3.4.3 独自性得点

独自性得点は AI 採点と人間採点のいずれも有意ではなかった。

3.5 インタビュー

実験群の 2 名，統制群の 4 名を除くすべての参加者が，散歩中に何らかの MW をしたと回答していた。散歩中に最も長く MW した内容について，両群とも**現在**と答えた人が最も多かったが（図 4），実験群は散歩自体について，統制群は妨害課題についての内容が多く，群間で内容は異なっていた。

図 4　最も多く MW した内容



4.　考察

本研究では，散歩中に生じる MW が散歩前後の気分や創造性にどのような影響を及ぼすのか，実験的手法によって検討した。

4.1 散歩 MW の気分への効果

まず，散歩中の MW が気分に及ぼす影響について,散歩後にポジティブ気分が上昇し，ネガティブ気分が低下したことから，散歩によって気分がポジティブに改善されることが示された。この結果は，散歩自体にポジティブな効果があった可能性を示唆している。また，散歩中の MW を比較したところ，意図的 MW では時期の固定効果が有意であったことが示された。実験群は孵化効果の教示，統制群は妨害課題として回答したクイズのことを鑑みると，両方の介入が意図的な MW を高めた可能性，あるいは散歩自体に意図的 MW を促進する効果があった可能性が考えられる。また，インタビュー結果からも，同じ散歩であっても実験群と統制群では MW の内容やタイプが異なっていた可能性がある。

4.2 散歩 MW の創造性への効果

AUT については，流暢性は散歩後にアイデア回答数が全体的に増加したものの，群間の有意差は見られなかった。柔軟性は，孵化効果の教示がなかった統制群の方が散歩後の得点が高くなり，仮説と反対の結果が示された。実験群の柔軟性得点が上がらなかった理由について，孵化効果を促す教示が認知的負荷になり，参加者が散歩後に良いアイデアを出さなければならないといった緊張や不安などが生まれたおそれがある。本研究では統制群にのみ，妨害課題によって自由な MW

を統制することを意図したが，実験群でも教示によって異なる種類の認知的負荷がかかったおそれがある。あるいは実験群における，ひらめきを得られるように思考を彷徨わせるという教示自体が人によっては困難であった可能性も否めない。実際に図 2 から，実験群の参加者は MW の変化量に個人差があったことが分かる。また，実験群への教示は，一般的に想定される孵化効果，すなわち潜在的な意識によるひらめきを促すものではなく，むしろ顕在的に「考えようとする」操作となっていたことが挙げられる。一方で，統制群に課された妨害課題は，創造性課題から離れる時間を生み出し，結果として孵化効果が生じていた可能性が考えられる。さらに，独自性はいずれも有意な差は見られなかったが，回答がどれほど独創的かを 5 段階で評定する人間採点と AI 採点とでは採点の意味合いが違うことが考えられる。創造的自己効力感もいずれも有意な差は見られなかったが，先述したように創造性課題は実験群への介入の効果が見られなかったことから，参加者自身が明確な創造性の向上を感じなかった可能性もある。

4.4 本研究の限界と展望

最後に本研究の限界について述べる。第一に，散歩中の思考を尋ねる方法である。本研究では，インタビューは山岡・湯川（2019）の先行研究の思考プローブの分類をもとにしたが，散歩中の思考は次々と変化するため，実験室内での統制された環境で瞬時に思考内容を測定する思考プローブの分類を，そのまま本研究へと当てはめることには限界があった。また，本実験では散歩中の思考について，回顧的報告を求めたため，記憶の誤謬や修正が含まれた可能性があったという課題も挙げられる。そのため，今後はリアルタイムの思考サンプリングや経験サンプリングによる測定を検討したい。

第二に，散歩中の MW の意図性と非意図性の区別である。回顧的な回答では，散歩中の MW を意図的か非意図的か分けることに限界があった。本研究では，孵化効果の教示をして自由な MW を誘発する実験群と，自由な MW を統制する統制群に群分けをしたが，実験群の 2 名，統制群の 4 名を除くすべての参加者が散歩中に何らかの MW をしたと回答していたために，教示や妨害課題によって MW を十分に統制できていなかったと考えられる。今後は意図的・非意図的 MW を統制する手法や思考内容をコントロールするための訓練手法の開発などの工夫が必要である。

散歩はポジティブで自由な MW を誘発しやすく，こうしたポジティブな MW は創造性や精神的健康の向上に寄与する可能性がある。したがって，本研究での課題を踏まえ改善した上で，実施方法によっては教育場面でのアイデア想起の取り組みや，臨床場面での気分の改善を促す介入として有用であることが考えられる。

文献

Bar, M (著). 横澤一彦 (訳). (2023)マインドワンダリング: さまよう心が育む創造性. 勁草書房.

DeepL Translation (https://www.deepl.com/ja/translator). 2026 年 7 月 9 日最終閲覧.

石黒千晶, 清水大地, & 清河幸子. (2022). 「創造的自己」 から創造性研究を捉え直す. *認知科学, 29*(2), 289-292. https://doi.org/10.11225/cs.2022.021

石黒千晶, 山川真由, & 櫃割仁平. (2025). 拡散的思考課題の自動採点に関するシステマティックレビュー. *認知科学, 32*(1), 26-40. https://doi.org/10.11225/cs.2025.052

小塩真司, & 阿部晋吾. (2012). 日本語版 Ten Item Personality Inventory (TIPI-J) 作成の試み. *パーソナリティ研究, 21*(1), 40-52. https://doi.org/10.2132/personality.21.40

佐藤徳, & 安田朝子. (2001). 日本語版 PANAS の作成. *性格心理学研究, 9*(2), 138-139. https://doi.org/10.2132/jjpjspp.9.2_138

Raichle, M. E. (2015). The brain's default mode network. *Annual review of neuroscience, 38*(1), 433-447. https://doi.org/10.1146/annurev-neuro-071013-014030

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual review of psychology, 66*(1), 487-518. https://doi.org/10.1146/annurev-psych-010814-015331

Thabane A, Arora V, Boilard J, Sutoski A, Parpia S, Calic G, et al. (2026).The impact of walking on creative thinking:A systematic review and meta-analysis. *PLoS One 21*(5) : e0347878. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0347878

The Creativity Assessment Platform. Automated test delivery and AI scoring. (https://cap.ist.psu.edu). 2026 年 7 月 9 日最終閲覧.

山岡明奈, & 湯川進太郎. (2016). マインドワンダリングが創造的な問題解決を増進する. *心理学研究, 87*(5), 506-512. https://doi.org/10.4992/jjpsy.87.15057

山岡明奈, & 湯川進太郎. (2019). 日本語版意図的/非意図的 MW 傾向尺度の作成と信頼性・妥当性の検討. *教育心理学研究, 67*(2), 118-131. https://doi.org/10.5926/jjep.67.118

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).

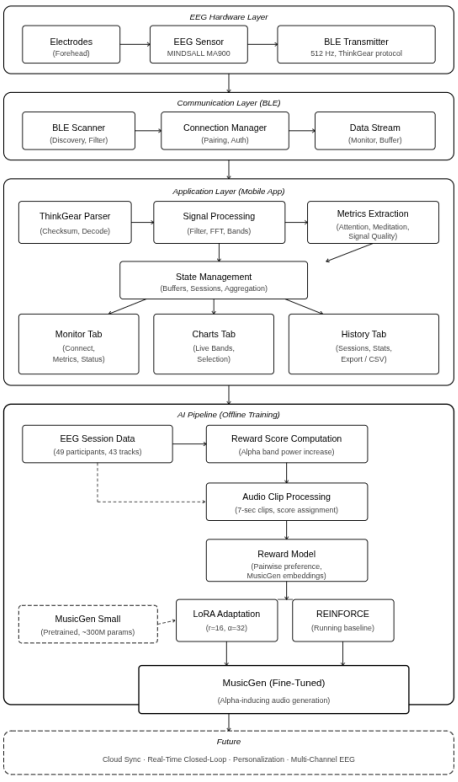
MusicGen encodes raw audio via EnCodec into discrete token sequences, then models them autoregressively with a transformer conditioned on text descriptions; its 300M-parameter "small" variant is computationally tractable for RL fine-tuning, which is why we adopt it as the base model here rather than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

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

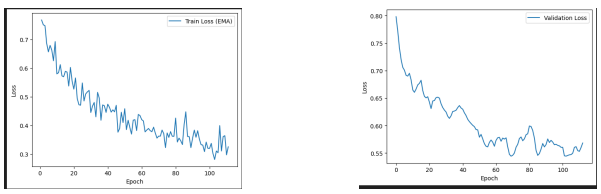
where $A_t = r - b_t$ is the advantage over an exponentially smoothed baseline (α_b=0.6) and λ=0.005 is the entropy regularization coefficient. Sequences of 256 tokens (≈5 s) are sampled at temperature 0.8, top-k 128, with gradient accumulation over 2 steps for 500 iterations, saving checkpoints every 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.