Exploring the Impact of Music on Brain Activity during Testing: An EEG-based Investigation

Dhyego Tavares
Federal University of Bahia
Salvador, Bahia, Brazil
dhyegocruz@ufba.br

Roselane Silva Farias
Federal University of Bahia
Salvador, Bahia, Brazil
silva.roselane@ufba.br

Michelle Larissa
Federal University of Bahia
Salvador, Bahia, Brazil
michellellcarval@gmail.com

José Garcia Vivas Miranda
Federal University of Bahia
Salvador, Bahia, Brazil
vivasm@gmail.com

Eduardo Santana de Almeida
Federal University of Bahia
Salvador, Bahia, Brazil
esa@rise.com.br

Abstract

Software testing is a cognitively demanding task in which human factors such as attention, stress, and fatigue critically influence tester performance. While music is commonly used by professionals to enhance focus, its effects on the brain activity of software testers remain unexplored. This study investigates the influence of instrumental LoFi music on the cognitive performance and neural activity of professional software testers through a controlled EEG experiment. Fourteen experienced testers were divided into music and control groups and performed 12 testing tasks spanning code comprehension, syntax error detection, logic error identification, and test case creation. EEG data were recorded using a 64-channel cap and analyzed for power spectral density and event-related desynchronization/synchronization (ERD/ERS) across Delta, Theta, Alpha, and Beta frequency bands. The Music group exhibited significantly lower frontal Delta power across all task blocks ($d = -1.28$ to -1.69, $p \leq .033$), indicating sustained neural readiness, and significant Theta differences in mid-experiment tasks ($p \leq .044$). Alpha and Beta bands showed large directional effect sizes ($d \approx -0.81$ to -0.88) consistent with a flow-like attentional state, though underpowered for formal confirmation. Behaviorally, the Music group outperformed controls in comprehension (87.5% vs. 79.17%), syntax error detection (66.67% vs. 54.17%), and logic error identification (45.83% vs. 33.33%). In test case creation, however, the Music group scored below controls (66.67% vs. 70.83%), a result neurophysiologically explained by Alpha synchronization indicating attentional drift during generative reasoning.

Keywords

EEG, Experiment, Music, Software Testing, Brain.

1 Introduction

Software testing is a fundamental pillar of software engineering, ensuring that systems meet stringent quality and reliability standards [31]. Its primary goal is to verify that software processes, methodologies, and deliverables adhere to established requirements [44], while assessing functional correctness, stability, and robustness under diverse operational conditions [16, 46]. By embedding rigorous evaluation practices throughout development, testing supports engineering decisions and the delivery of reliable and deployable software [31, 56].

Despite its critical role, software defects remain a persistent challenge, with defect density imposing substantial costs on both the software industry and society [33]. By exploring the cognitive processes of software testers during testing activities, targeted strategies can be developed to better support their work, enhancing performance and improving overall software reliability.

As testing environments become increasingly complex, there is growing interest in the human factors that influence testing effectiveness. Cognitive states such as attention, stress, and fatigue can significantly impact a tester's performance, presenting a critical challenge for both research and practice. Given the meticulous and cognitively demanding nature of testing, maintaining high levels of concentration and productivity is essential.

One potential approach to supporting focus in such demanding environments is the use of background music. Many professionals, including software developers, routinely listen to music while working [3, 14]. Although the emotional and cognitive effects of music have been explored in previous studies [19, 35], this topic remains underexplored in the specific context of software testing.

Understanding how music influences software testers' performance can provide valuable insights into potential strategies for supporting their work. In particular, music may serve as a cognitive aid to enhance testing efficiency—an increasingly relevant need given the growing complexity of modern software systems and the demand for more accurate and reliable testing processes [18]. Moreover, as a fundamental aspect of human cognition, music provides rich sensory, cognitive, emotional, and social stimulation [55]. Its documented ability to regulate emotions, enhance mood, and reduce stress [2, 5, 24, 27, 28] suggests that it may positively influence both the emotional state and concentration of software testers. However, its effects on performance in software testing tasks remain largely unexplored, highlighting the need for empirical investigation.

To better understand this potential, it is essential to examine not only the behavioral effects of music but also the underlying cognitive processes and brain activity of testers during testing tasks. Current research still lacks insight into how testers perceive and analyze code, the cognitive load associated with these activities, and how external factors such as music may influence these processes. Addressing this gap aligns with the growing interest in neuroscience within software engineering, where techniques such as EEG and functional magnetic resonance imaging (fMRI) have been increasingly adopted to study developers' cognitive states [10, 25, 26, 29, 39–41, 50, 51, 54].

Prior studies have made important advances by investigating activities such as code comprehension, cognitive load, programming tasks, and differences across experience levels [8, 10, 11, 13, 20, 26, 27, 29, 39, 54, 66]. The findings indicate, for instance, that experienced developers exhibit enhanced memory-related brain activation [10, 20, 29, 54] and lower cognitive workload during development tasks [10].

In this paper, we investigate the influence of music on the cognitive processes and brain activity of experienced software testers during testing tasks, through a controlled experiment with 14 participants, each with at least three years of experience. By analyzing both neural and behavioral data, we aim to understand how testing activities affect cognitive functions and how music may modulate these effects. To guide our investigation, we address the following research questions:

RQ1: What are the effects of music on the brain activity patterns of software testers during testing tasks?

This research question investigates whether listening to music influences neural activity during software testing. We analyze EEG data and cognitive load indicators to identify significant differences in brain activity patterns between music and non-music conditions.

RQ2: How does listening to music influence the performance of software testers across different testing tasks?

Different testing activities impose varying cognitive demands, which may interact differently with external stimuli such as music. By examining performance across tasks of distinct nature, such as code comprehension, error detection, and test case writing, we aim to understand whether the effects of music are consistent or task-dependent. This analysis enables us to identify how music may enhance or hinder performance depending on the cognitive requirements of each task, providing a more nuanced understanding of its role as a potential cognitive aid in software testing.

Our results indicate that participants exposed to music exhibited more stable neural patterns, particularly in the Beta and Delta bands, which are commonly associated with sustained attention and cognitive engagement [22, 47]. In addition, we observed event-related desynchronization (ERD) [42] in the Theta, Alpha, and Beta bands, suggesting increased cortical activation and task-related cognitive processing. These neural patterns were accompanied by improved task accuracy, supporting the potential of music as a non-invasive aid to enhance focus and performance in software testing.

In summary, our study makes the following contributions:

- To the best of our knowledge, this is the first study to employ EEG to investigate brain activity in software testers across a range of testing tasks.
- We designed a set of 12 testing tasks, including test automation code comprehension, syntax and logic error detection, and test case writing, providing reusable resources for future research.
- We provide all the necessary artifacts for replicating this study (see Section *Artifacts Availability*).

2 Background and Related Work

This study lies at the intersection of software engineering and neuroscience, requiring an integrated view of how human cognition influences software testing tasks. In this section, we review three key areas that ground our work. First, we introduce electroencephalography (EEG) as a method for capturing cognitive processes through brain activity signals. Second, we discuss prior EEG-based studies in software engineering, highlighting insights into cognitive load, expertise, and task-dependent neural responses. Finally, we examine research on the cognitive and emotional effects of music, with a focus on its potential role in modulating attention, stress, and performance. Together, these perspectives motivate our investigation of music as a contextual factor influencing software testers' cognitive states and behavior.

2.1 Electroencephalography in Cognitive Analysis

EEG is a widely used non-invasive technique for measuring brain activity through electrical signals generated by neuronal firing [47]. Its high temporal resolution and relative cost-effectiveness make it particularly suitable for studying dynamic cognitive processes in software engineering tasks [61].

EEG captures oscillatory activity across distinct frequency bands, commonly associated with different cognitive states. Delta (0.5–4 Hz) has been linked to cognitive load in alert individuals, Theta (4–8 Hz) to working memory and sustained attention, Alpha (8–13 Hz) to relaxed wakefulness and its suppression during cognitive engagement, and Beta (13–30 Hz) to active thinking and focused problem-solving. Table 1 summarizes these bands and their associated brain states.

Table 1: EEG Frequencies and Associated Brain States

EEG Frequency	Associated State of Brain
Delta (0.5 - 4Hz)	Deep sleep
Theta (4 - 8Hz)	Creativity / Memory / Light Sleep
Alpha (8 - 13Hz)	Relaxed mental activity / No Concentration
Beta (13 - 30Hz)	Focused / High Mental Activity

These frequency-based markers provide a foundation for interpreting cognitive effort, attention, and mental workload during software-related activities, making EEG a valuable tool for investigating how external factors—such as music—may influence tester cognition.

2.2 EEG in Software Engineering

Research at the intersection of software engineering and neuroscience has increasingly adopted EEG to investigate cognitive processes during development tasks. Compared to fMRI, EEG offers a practical balance between temporal resolution, cost, and experimental flexibility [17, 34, 61].

Early studies primarily focused on novice programmers performing code comprehension tasks. For example, Calcagno et al. [8] observed increased Delta, Theta, and Beta activity during code analysis, associating these patterns with working memory engagement and cognitive effort. Subsequent work has examined differences across expertise levels, showing that experienced developers tend to exhibit lower cognitive load and more efficient neural patterns [29, 39].

More recent research highlights that neural responses vary significantly depending on task characteristics. Tasks requiring deeper reasoning, such as logic understanding, tend to elicit stronger Theta and Beta engagement compared to syntactic reading [26, 66]. These findings reinforce that cognitive processes in software engineering are highly task-dependent. Despite these advances, EEG-based studies focusing specifically on software testers remain scarce, motivating the present investigation.

2.3 Music and Cognitive Processes

Music plays a central role in human cognition, particularly in the regulation of emotions, attention, and mental effort [55]. Its ability to evoke affective responses makes it a valuable lens for studying the interaction between emotional and cognitive processes, as well as their neural correlates [24]. Recent advances in neuroscience have further highlighted the mechanisms underlying music-induced experiences, showing that neural synchronization—especially between frontal and temporal regions in the theta band—is associated with emotional engagement and pleasurable responses to music [60].

Beyond its emotional effects, music has been widely investigated as a contextual factor influencing performance and productivity in work environments. Early studies explored its role in industrial and routine tasks, suggesting that music can reduce perceived workload and improve job satisfaction [36, 37, 57]. In knowledge-intensive domains, such as software engineering, more recent work indicates that professionals frequently use music to regulate mood, maintain focus, and cope with stress during cognitively demanding activities [3].

From a cognitive perspective, music may influence performance through multiple pathways. On one hand, it can enhance sustained attention, stabilize emotional states, and reduce mental fatigue, potentially supporting task execution. On the other hand, music may also act as a competing stimulus, particularly in tasks that require complex reasoning or generative thinking, where additional cognitive resources are needed.

3 Experimental Study

This section describes the design of our empirical study. We first present the participant recruitment process, followed by the experimental design, including tasks, tools, and procedures. Finally, we detail the data acquisition and analysis methods.

3.1 Participants

We conducted a controlled experimental study with 14 software testers who performed testing tasks while undergoing EEG recording. The study was approved by the Research Ethics Committee of School of Nursing, Federal University of Bahia - UFBA, in May 2024 (Approval No. 78645624.2.0000.5531).

Participants were recruited through convenience sampling [64] from local companies. All participants had professional experience in test automation using Java, as well as prior exposure to Selenium and XPath, which served as inclusion criteria.

Table 2 summarizes participants' demographic characteristics. The sample included professionals with varying levels of expertise, providing a realistic representation of industry testers.

Table 2: Demographic Information of the Participants

Demographic Variables	
Gender	**Participants**
Men	10
Women	4
Expertise Level	
Novice (0-3 years as a tester)	6
Intermediate (4-9 years as a tester)	5
Advanced (10+ years as a tester)	3
Experience	**Mean (Median, SD)**
Years of Programming Experience	9 (7, 5.3)
Years of Java Experience	6.6 (6.0, 5.1)
Years of Testing Experience	6.1 (5.0, 4.4)

Participants were randomly assigned to two groups of equal size (n=7). The control group performed all tasks without music, while the experimental group completed the same tasks while continuously listening to a standardized LoFi playlist via headphones.

To reduce confounding effects, we excluded individuals using psychotropic medication or presenting conditions that could affect cognitive processing, such as neurological impairments or significant visual deficiencies.

3.2 Choice of Musical Stimuli

Previous studies have shown that music can enhance productivity; however, certain types—particularly those with lyrics or highly danceable rhythms—may instead divert attention from the task at hand [3, 27, 28, 37]. Instrumental genres such as LoFi, classical, or ambient music are often considered less distracting and more suitable for supporting focus. While music preference is inherently subjective, we selected instrumental LoFi tracks for this study, as they typically lack vocals or lyrical content [63]. Specifically, four albums from *LoFi Girl* were chosen as the auditory stimulus (available in the supplementary material, see Section *Artifacts Availability*).

To ensure methodological rigor, all participants listened to the same LoFi playlist. Allowing personal music choices could introduce confounding variables—such as genre, emotional associations, and individual preferences—making it more difficult to isolate the effects of music on cognitive performance and brain activity. Standardizing the musical stimulus improved experimental control, EEG data comparability, and internal validity.

By standardizing the stimulus, we aimed to isolate the physiological impact of music from idiosyncratic emotional or cognitive reactions associated with personal preferences. Furthermore, to verify the suitability of this choice, brief post-experiment interviews were conducted; none of the participants reported an aversion to the selected LoFi tracks, suggesting that the stimulus was generally well tolerated and did not introduce significant negative bias.

3.3 Experimental Design

The experiment consisted of two phases: a baseline phase and a task execution phase. In EEG experiments, the baseline phase serves as a reference point for detecting changes in brain activity and as a benchmark for comparing task-related EEG data. During the baseline phase, participants remained seated with eyes open for three minutes while focusing on a fixation cross.

In the task phase, participants performed four blocks of testing activities (see Figure 1). Each block corresponded to a distinct testing task. A 30-second rest period was included between tasks to minimize cognitive carryover effects. Participants in the music group listened to the LoFi playlist continuously throughout both phases, while the control group performed the experiment under silent conditions.

Each experimental session, comprising the baseline phase and all four task blocks, was completed in a single sitting on the same day, with a total duration of approximately 35 minutes per participant. Sessions were scheduled during working hours (morning or afternoon, depending on participant availability), and both groups were distributed across similar time slots to minimize potential confounding effects related to circadian variation in alertness.

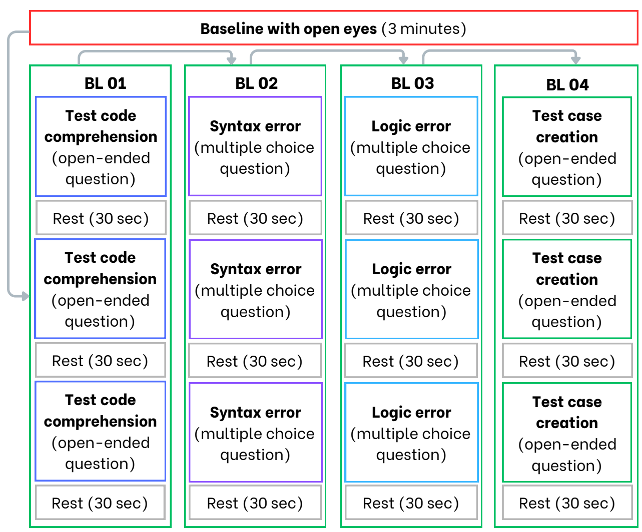


Figure 1: Visualization of the experimental design, illustrating the four task blocks.

3.4 Experimental Tool

We utilized PROPHET[1], a Java-based infrastructure designed for program comprehension experiments. This tool has been widely used in empirical studies since 2012 [12, 21, 50, 52, 53].

Java was selected as the programming language for several reasons. It is widely preferred in test automation due to its robust features and frequent adoption in program comprehension research [6]. Additionally, Java is among the most commonly used languages in code comprehension experiments [65], and prior research on program comprehension frequently includes Java code snippets [7, 38, 51, 54].

[1]The source code is available at https://github.com/feigensp/Prophet

3.5 Testing Tasks

We designed four types of testing tasks to reflect common software testing activities.

BL01 – Test Code Comprehension: This block comprised three open-ended questions aimed at assessing participants' understanding of test code. Participants were provided with a code snippet written in the designated programming language and were required to determine the test code's output. The objective of this block was to evaluate participants' ability to comprehend and interpret the provided code, simulating activities such as reviewing test logic and interpreting automated test results.

BL02 - Syntax error: This block included three multiple-choice questions centered on **finding syntax errors**. Participants were shown a snippet written in the specified programming language and asked to identify the line they believed contained a syntax error. The goal of this block was to evaluate participants' debugging skills and their ability to detect syntax issues in test code simulating activities such as static analysis, code review, and test code maintenance.

BL03 - Logic error: This block, similar to the previous one, comprised three multiple-choice questions focused on **finding logic errors**. Participants were required to identify the line they believed contained a logic error in the given snippet. The purpose of this block was to assess participants' ability to debug test code and identify logical flaws, simulating activities such as fault localization and test behavior validation.

For instance, Listing 1 presents a sample snippet that demonstrates the interaction with a dynamic list of web elements using Selenium in Java. In this snippet, the error occurs on **line 6**, where the *xpath* expression is incorrectly formatted, which may lead to runtime issues. Additionally, the loop logic may cause an *IndexOutOfBoundsException* due to the incorrect starting index (i = 1 instead of i = 0).

```
1  public class LogicError {
2
3      List<WebElement> ListofItems = wd.findElements(By.xpath("//*[
           starts-with(@id,'result_')]//div//div[1]//div//a//img"));
4
5      for (int i = 1; i <= ListofItems.size(); i++) {
6          ListofItems = wd.findElements(By.xpath("//*[starts-with(
               @id,'result_')]//div//div[1]//div//a//img"));
7
8          ListofItems.get(i).click();
9          wd.manage().timeouts().implicitlyWait(30, TimeUnit.SECONDS
               );
10         System.out.print(i + " element clicked\t--");
11         System.out.println("pass");
12         wd.navigate().back();
13     }
14 }
```

Listing 1: Test code in Selenium/Java with syntax error

BL04 - Test Case Creation: This block consisted of three questions centered on **test case writing**. Participants were tasked with drafting test scenarios for a given problem and extending a test case in the specified programming language. The objective of this block was to evaluate participants' proficiency in creating test cases in natural language and their expertise in test automation through maintenance-related tasks, including designing new tests and refining existing test suites.

3.6 Data Acquisition and Analysis

The data collection process followed standard neurophysiological protocols recommended by the International Federation of Clinical Neurophysiology (IFCN). To ensure methodological rigor and data integrity, two researchers supervised each session. One researcher maintained a detailed logbook to document any behavioral or technical anomalies, such as involuntary participant movements, sneezes, or transient equipment malfunctions. Simultaneously, the second researcher monitored EEG signals in real time to manage the experimental flow and ensure consistent signal quality. Anomalies recorded in the logbook were cross-referenced during preprocessing; segments corresponding to significant behavioral artifacts or hardware instability were flagged for manual inspection and subsequent exclusion to prevent data contamination.

Brain electrical activity was recorded using a Compumedics Neuroscan Neuvo 64-channel EEG system, with the EEG cap configured according to the international 10-10 system, which provides standardized electrode placement across the scalp. The Cz electrode served as the reference. To optimize signal conduction, electrode-skin impedance was strictly maintained below 50 kΩ. Data were sampled at 1000 Hz, ensuring high temporal resolution for analyzing fast-changing neural dynamics. All recordings were captured using *Profusion*[2] software.

To differentiate between cognitive states, we employed an ensemble approach rather than relying solely on raw EEG traces. This methodology integrated five techniques: (i) task-based baseline normalization, (ii) frequency-specific spectral markers, (iii) cortical topography, (iv) event-related dynamics (ERD/ERS), and (v) convergence with behavioral performance. This integration reduces ambiguity when different cognitive states might produce superficially similar EEG patterns.

Preprocessing was conducted using EEGLAB in MATLAB R2018b. The data were band-pass filtered between 0.5 and 50 Hz to isolate relevant neural components and to remove power-line noise. The continuous signal was segmented into 1.5-second epochs. To ensure a clean dataset, we applied Independent Component Analysis (ICA) to identify and remove physiological artifacts (eye blinks, cardiac signals, and muscle activity). Furthermore, epochs with amplitudes exceeding $\pm 100\ \mu$V were automatically rejected, as such values typically indicate non-neural interference [9]. Finally, the researchers performed a manual inspection, guided by the session logbook, to confirm the removal of any remaining noise.

Following artifact removal, Power Spectral Density (PSD), a measure of signal power across frequency bands, was estimated using Welch's method [62] on 1-second non-overlapping epochs. Absolute power was calculated for four frequency bands: Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), and Beta (13–30 Hz). For the Fast Fourier Transform (FFT) analysis, a Bonferroni correction was applied to account for multiple comparisons across the frequency bands, ensuring the statistical validity of the findings [62].

To assess the normality of the data, we applied the Shapiro–Wilk test [49], however, because N has a limited value, this test rarely rejects normality. To account for this uncertainty, we adopted a dual-test strategy: the Mann-Whitney U test [30] was applied as the primary nonparametric method, given its robustness to distributional assumptions, and was supplemented by an independent samples t-test [58] to increase sensitivity under the tentative assumption of normality.

[2] https://www.compumedics.com.au/en/products/profusion-eeg/

4 Results

This section presents the analysis of the processed data, aimed at addressing the research questions outlined in the study. The results are organized to highlight key patterns and statistically significant findings, offering insights into the cognitive effects observed under the experimental conditions.

4.1 RQ1: What Are the Effects of Music on the Brain Activity Patterns of Software Testers During Testing Tasks?

The analysis of brain activity amplitudes, segmented by cortical regions and frequency bands, reveals clear differences between the *Music* (Experimental) and *NoMusic* (Control) groups. As illustrated in Figure 2, in the *frontal* and *central* regions, the *NoMusic* group consistently exhibited higher Delta and Theta amplitudes—established biomarkers of increased working memory load and mental effort. This pattern suggests that testers in silence required greater neuronal recruitment to perform the same tasks. In contrast, the reduced amplitudes observed in the *Music* group indicate that music may have facilitated information processing with lower cognitive expenditure, consistent with the concept of neural efficiency.

The Alpha band exhibited smaller and more stable amplitudes in the *Music* group, suggesting more consistent cortical engagement, whereas the *NoMusic* group showed greater oscillations between states of focus and mind-wandering. Regarding the Beta band, the *NoMusic* group presented higher amplitudes, particularly in the parietal and temporal regions—a pattern often associated with stress and mental fatigue under pressure. In contrast, the *Music* group maintained lower and more controlled Beta levels, indicative of a flow-like state characterized by high precision and reduced psychophysiological effort. Moreover, the temporal and occipital regions showed no signs of sensory congestion in the *Music* group, despite the additional auditory stimulus, suggesting that music functioned as an anxiety regulator rather than a distractor, thereby optimizing cortical resource allocation.

4.1.1 ***Event-Related Desynchronization and Synchronization (ERD/ERS)***. Event-Related Desynchronization and Event-Related Synchronization [42] are widely used metrics in EEG research to assess dynamic changes in the power of brain oscillations in response to specific events, such as sensory stimuli, cognitive demands, or motor actions [8, 15, 23, 48, 59].

To quantify these changes, ERD/ERS is typically calculated as the percentage of power variation between a reference (baseline) interval and a task-related interval. The calculation for a single electrode is defined in Equation 1, while the average for a specific cortical region is defined in Equation 2, following the methodology proposed by Pfurtscheller and Lopes Da Silva [43]:

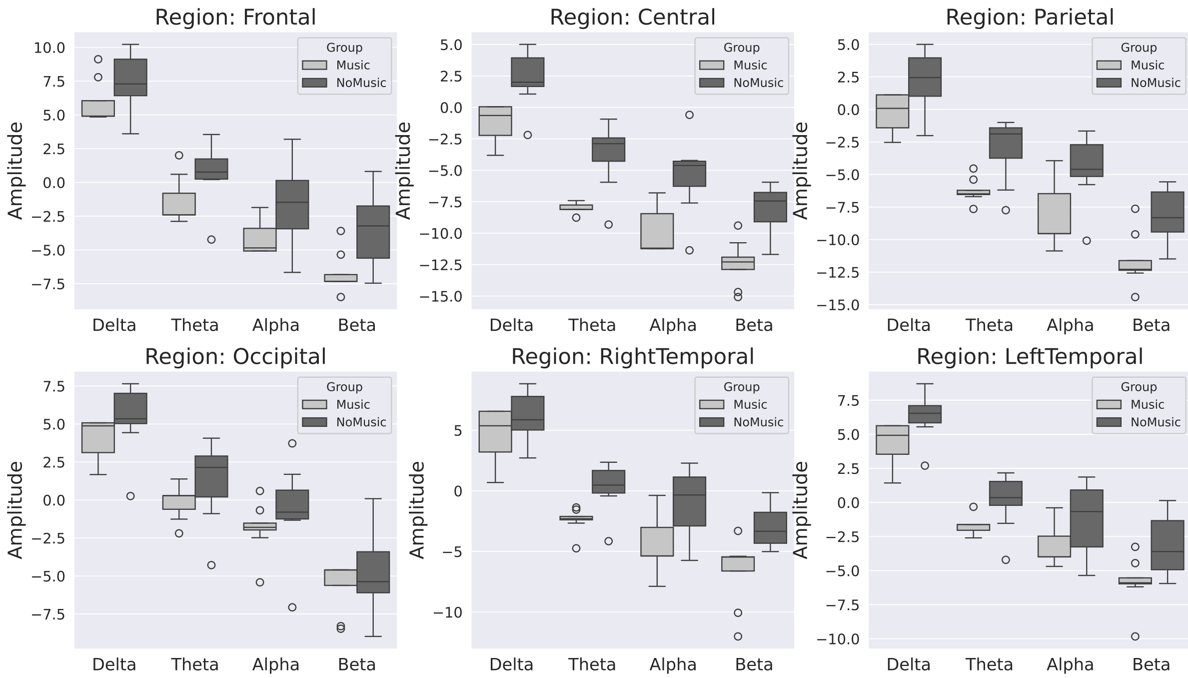


Figure 2: Box plot showing the distribution of EEG Band amplitudes across different brain regions for both control and experimental groups.

$$\mathrm{ERD/ERS}_{\mathrm{Electrode}}(\%) = \frac{P_{\mathrm{ref}} - P_{\mathrm{task}}}{P_{\mathrm{ref}}} \times 100 \tag{1}$$

$$ERD/ERS_{\mathrm{Region}}(\%) = \frac{\sum_{i=1}^{N} ERD/ERS_{\mathrm{electrode}_i}}{N} \tag{2}$$

Where:

- P_{ref} represents the average power of a given frequency band during the baseline period;
- P_{task} represents the power during the task or event-related interval;
- N represents the number of channels (electrodes) in the region.

Negative ERD/ERS values indicate a task-related decrease in band power relative to the baseline and are interpreted as **ERD**, reflecting reduced oscillatory activity. Conversely, positive values indicate a task-related increase in band power compared to the baseline and are interpreted as **event-related synchronization (ERS)**, reflecting enhanced oscillatory activity [23, 43].

Figure 3 presents the average ERD/ERS (%) values grouped by frequency band (Beta, Alpha, Theta, and Delta) and experimental condition (with and without music), and organized by testing task (BL01, BL02, BL03, and BL04) in the frontal cortex, a region associated with key cognitive processes.

ERD/ERS at BL01: In Figure 3 (a) The Music group showed stronger Alpha ERD than the NoMusic group, suggesting greater concentration, while Theta remained similarly low in both groups, indicating reduced working memory load. Beta desynchronization was more pronounced in the NoMusic group.

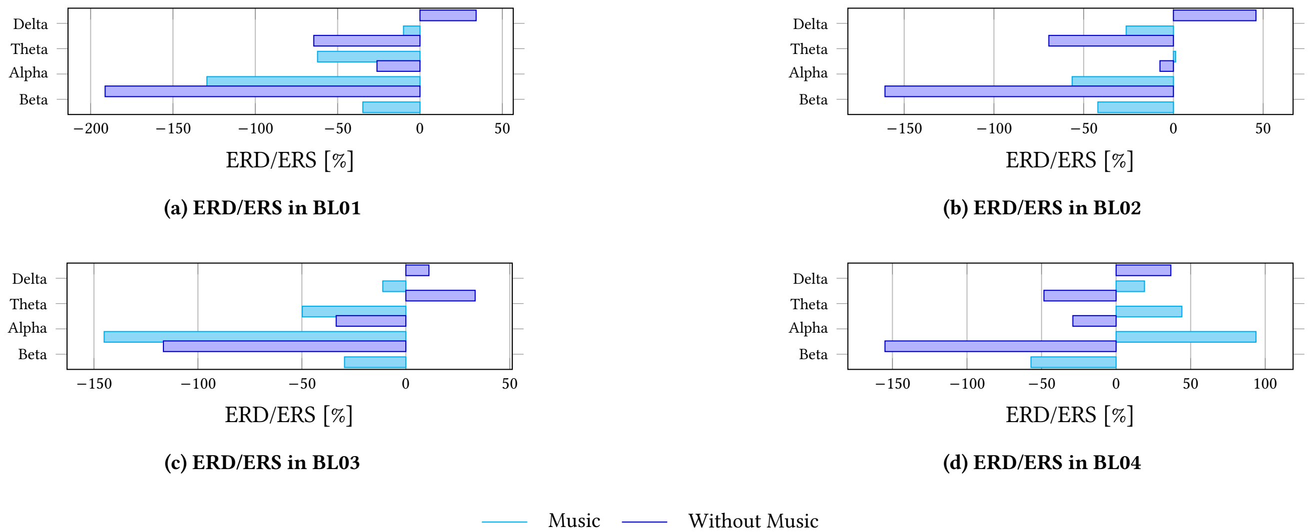
ERD/ERS at BL02: As shown in Figure 3 (b), Alpha ERD was markedly stronger in the Music group. Theta showed a divergent pattern—strong desynchronization in the NoMusic group versus slight synchronization (ERS) in the Music group, suggesting comparatively higher working memory engagement under music. Beta desynchronization was again stronger in the NoMusic group.

ERD/ERS at BL03: As shown in Figure 3 (c), Alpha ERD was more pronounced in the Music group, though a similar pattern was observed in the NoMusic group at lower magnitude. Theta showed a divergent pattern—synchronization (ERS) in the Music group versus strong desynchronization in the NoMusic group, indicating comparatively greater working memory activity under music. Beta desynchronization was again less pronounced in the Music group.

ERD/ERS at BL04: As shown in Figure 3 (d), this block shows the sharpest contrast between conditions: the Music group shifted from Alpha ERD (seen in BL01–BL03) to Alpha ERS, suggesting reduced concentration, while the NoMusic group maintained ERD. Theta also inverted, with the Music group now showing synchronization and the NoMusic group desynchronization. Beta followed the same directional pattern observed in prior blocks.

4.1.2 ***Statistical Analysis of Power Spectral Density.***

To complement the ERD/ERS analysis, we conducted a formal statistical comparison of Power Spectral Density (PSD) values between the Music and NoMusic groups across all task blocks (BL01-BL04) for each frequency band in the frontal lobe. Before the inferential analysis, normality was assessed using the Shapiro-Wilk test [49]; however, given the small sample size ($n = 7$ per group), this test has



(a) ERD/ERS in BL01

(b) ERD/ERS in BL02

(c) ERD/ERS in BL03

(d) ERD/ERS in BL04

Figure 3: Mean ERD/ERS (%) by Frequency Band and Experimental Condition (with and without music) across Testing Tasks in the Frontal Cortex.

Table 3: Mann-Whitney U Test and Independent Samples t-Test Results (Music vs. NoMusic)

		Mann-Whitney U				t-Test		
Band	**Task**	**U**	**Z**	**p**	**r**	**t**	**p**	**d**
Alpha	BL01	19.0	−0.703	0.535	0.188	−0.670	0.516	−0.358
	BL02	14.0	−1.342	0.209	0.359	−1.638	0.127	−0.876
	BL03	16.0	−1.086	0.318	0.290	−1.259	0.232	−0.673
	BL04	16.0	−1.086	0.318	0.290	−1.093	0.296	−0.584
Beta	BL01	19.0	−0.703	0.535	0.188	−0.772	0.455	−0.413
	BL02	14.0	−1.342	0.209	0.359	−1.517	0.155	−0.811
	BL03	17.0	−0.958	0.383	0.256	−0.862	0.406	−0.461
	BL04	20.0	−0.575	0.620	0.154	−0.469	0.647	−0.251
Delta	BL01	7.0	−2.236	0.026	0.598	−2.635	0.022	−1.408
	BL02	5.0	−2.492	0.011	0.666	−2.658	0.021	−1.421
	BL03	5.0	−2.492	0.011	0.666	−3.163	0.008	−1.691
	BL04	9.0	−1.981	0.053	0.529	−2.403	0.033	−1.284
Theta	BL01	11.0	−1.725	0.097	0.461	−1.913	0.080	−1.023
	BL02	8.0	−2.108	0.038	0.563	−2.254	0.044	−1.205
	BL03	10.0	−1.853	0.073	0.495	−2.317	0.039	−1.238
	BL04	9.0	−1.981	0.053	0.529	−2.037	0.064	−1.089

Effect: Sm. r<.30/d<.50; Med. r<.50/d<.80; Lg. r≥.50/d≥.80

limited statistical power and its results cannot be considered conclusive, and a failure to reject normality does not reliably confirm a normal distribution under such conditions [45]. To account for this uncertainty, we adopted a dual-test strategy: the Mann-Whitney U test [30] was applied as the primary nonparametric method, given its robustness to distributional assumptions, and was supplemented by an independent samples t-test [58] (with Levene's test for homogeneity of variances) to increase sensitivity under the tentative assumption of normality. Convergence between both tests was used as a criterion for greater confidence in the findings. Full results are presented in Table 3, with effect sizes reported as $r = Z/\sqrt{N}$ and Cohen's d, respectively.

Delta Band: The Delta band yielded the most robust and consistent between-group differences. As shown in Table 3, the Mann-Whitney U test identified significant effects in BL01, BL02, and BL03 (all $p \leq .026$, $r \geq .598$), with BL04 approaching significance ($p = .053$, $r = .529$). The t-test extended these findings to all four task blocks (all $p \leq .033$, $|d| \geq 1.28$). The consistently negative direction of U, Z, and d across all blocks indicates that the Music group exhibited systematically lower frontal Delta PSD throughout the experiment. Since the elevated frontal Delta in alert adults reflects reduced cortical engagement [4], this reduction is interpreted as evidence of greater neural activation and cognitive readiness in the Music group.

Theta Band Significant differences emerged in the mid-experiment blocks. The Mann-Whitney test detected a significant effect in BL02 ($p = .038$, $r = .563$), while the t-test additionally reached significance in BL03 ($p = .039$, $d = -1.24$), as detailed in Table 3. Although BL01 and BL04 did not reach significance, both presented large effect sizes ($|d| \approx 1.02$–1.09), suggesting that the limited sample size may have constrained statistical power.

Alpha and Beta Bands: Neither band reached statistical significance across any task block (all $p > .12$; see Table 3). Nevertheless, Cohen's d values were large for Alpha BL02 ($d = -0.88$) and Beta BL02 ($d = -0.81$), indicating potentially meaningful effects that the current sample size was insufficient to confirm. These results are consistent with the ERD/ERS patterns described in Subsection 4.1.1, where directional differences in Alpha and Beta were observed, but showed greater inter-participant variability.

4.2 RQ2: How Does Listening to Music Influence the Performance of Software Testers Across Different Testing Tasks?

To address this research question, the study analyzed the accuracy of participants' responses across four distinct testing tasks. As illustrated in Figure 4, participants in the music condition consistently achieved higher average accuracy in three out of the four

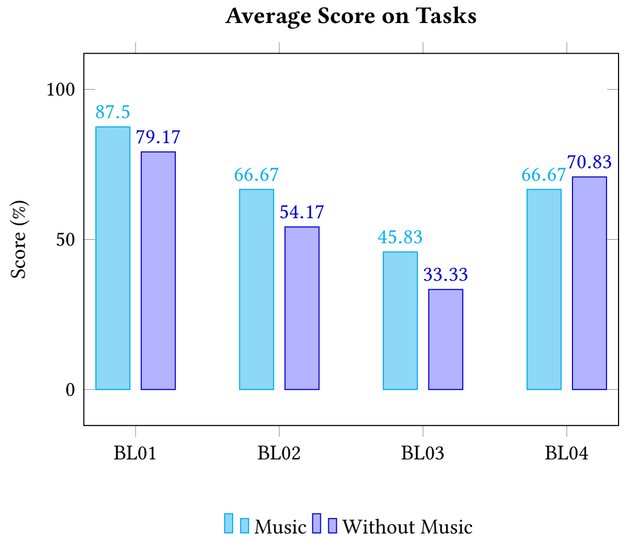


Figure 4: Average performance across the four software testing tasks.

task blocks (BL01, BL02, and BL03), which involved comprehension of test code, identification of syntax errors, and recognition of logic flaws, respectively. These results suggest that music may contribute positively to cognitive processing and problem-solving capabilities required in these specific testing activities. While these results are promising, additional studies are needed to strengthen this evidence.

However, in the test case creation task (BL04), participants who listened to music achieved a slightly lower average score compared to those in the control group (66.67% vs. 70.83%). Taken together, the findings support that music can enhance task performance by promoting cognitive stability and sustained attention—particularly in analytical tasks—while its effects may vary depending on the nature of the activity.

5 Discussion

Overall, our results indicate that music influenced both brain activity patterns and task performance, with effects varying across cognitive demands and testing activities.

The most robust neurophysiological finding concerns the Delta band, which showed significantly lower frontal power in the Music group across all task blocks ($d = -1.28$ to -1.69), indicating sustained neural readiness throughout the experiment. Theta reached significance in BL02 and, more tentatively, in BL03 (see Section 5.1 for a discussion of the discrepancy between the Mann-Whitney and t-test results in this block), while Alpha and Beta showed large directional effect sizes ($d \approx -0.81$ to -0.88) consistent with a flow-like attentional state, though underpowered for formal confirmation and thus interpreted here as exploratory rather than conclusive evidence. Behaviorally, the Music group outperformed controls in comprehension (+8.33 pp), syntax (+12.5 pp), and logic error detection (+12.5 pp). The exception, test case creation (BL04, −4.16 pp), is not an anomaly but a theoretically meaningful boundary condition: Alpha ERS in that block is directionally consistent with attentional drift precisely when open-ended generative reasoning was required, suggesting that music's benefits may be specific to convergent analytical tasks, a hypothesis that future work with larger samples should formally test.

The **Theta band** showed a more task-dependent pattern. Significant differences emerged in BL02 (Mann-Whitney: $p = .038$, $r = .563$) and BL03 (t-test: $p = .039$, $d = -1.24$), while BL01 and BL04, although non-significant, presented large effect sizes ($|d| \approx 1.02$–1.09). The ERD/ERS analysis adds interpretive depth in BL01 and BL02, where both groups showed Theta desynchronization (low working memory load), whereas in BL03, the Music group exhibited Theta ERS, a synchronization, indicating comparatively greater working memory engagement during the most cognitively demanding analytical task (logic error detection). This is consistent with BL03 being the task in which the Music group showed the largest absolute performance advantage (45.83% vs. 33.33%), and with Alpha ERD being most pronounced in the music condition, jointly suggesting a state of focused, memory-supported engagement.

For **Alpha** and **Beta** bands, no result reached statistical significance (all $p > .12$). However, this limitation should be interpreted in light of the effect sizes: Alpha BL02 ($d = -0.88$) and Beta BL02 ($d = -0.81$) both cross the conventional threshold for large effects, suggesting that the null results reflect insufficient statistical power ($n = 7$ per group) rather than the absence of true differences. Regarding ERD/ERS patterns, the Music group showed stronger Alpha desynchronization in BL01, BL02, and BL03, while the NoMusic group exhibited higher Beta amplitudes in parietal and temporal regions. These ERD/ERS patterns are descriptive and directionally consistent with, but do not statistically confirm, the interpretation of Alpha desynchronization as a marker of active cortical engagement and reduced idling, or of the observed Beta amplitude pattern as reflecting stress and psychophysiological load.

Together, these directional consistencies suggest that *music* promotes a flow-like cognitive state: high focus with low mental strain, even if the sample size prevented formal confirmation for these bands.

5.1 Task-Specific Modulation: From Neural Signal to Behavioral Outcome

The behavioral results in Figure 4 largely mirror the neurophysiological profile; however, this correspondence is not uniform across tasks, and this variability is itself theoretically meaningful.

In **BL01** (comprehension), the Music group outperformed the control group by 8.33 percentage points (87.5% vs. 79.17%). The ERD/ERS profile for BL01 shows pronounced Alpha ERD in the Music group, alongside comparable Theta levels between groups, suggesting that music enhanced attentional engagement without significantly increasing working memory demands. It is worth noting that, while Theta levels were formally compared between groups via Mann-Whitney/t-test, the Alpha ERD pattern itself was not subjected to statistical testing; its interpretation as reflecting attentional engagement is therefore descriptive and should be regarded as exploratory rather than confirmatory. With this caveat in mind, this represents a favorable configuration for comprehension tasks, which require sustained reading and inference, but not active generative processing.

In **BL02** (syntax error detection), the Music group showed the largest neural differentiation among the four blocks. Alpha ERD was markedly stronger, Beta desynchronization was more controlled, and Theta showed slight ERS under the music condition—accompanied by the largest effect sizes observed for non-Delta bands in the dataset ($d \approx -0.81$ to -0.88), although neither band reached statistical significance (Table 3). The behavioral advantage was 12.5 percentage points (66.67% vs. 54.17%). Taken together, these directional patterns are consistent with a neural profile of heightened focus and moderate working memory engagement, which would be well suited for pattern-matching tasks—though this interpretation remains exploratory given the lack of formal statistical confirmation for Alpha and Beta.

In **BL03** (logic error detection), despite being the most difficult task overall (scores of 45.83% and 33.33% respectively), the Music group maintained its advantage (+12.5 points). Uniquely in this block, the Music group exhibited Theta ERS, while the NoMusic group showed Theta ERD, an inversion that may indicate music helped participants preserve working memory resources that the control group struggled to maintain. It is worth noting that this Theta effect reached significance in the t-test ($p = .039$, $d = -1.24$) but not in the Mann-Whitney U test ($p = .073$), so this interpretation should be treated as tentative rather than conclusive. If confirmed by future work with larger samples, these findings would suggest that music may not only reduce cognitive load but also help preserve the cognitive capacity required for deeper analytical reasoning.

BL04 (test case creation) stands as the critical exception. In this task, the Music group scored 4.16 percentage points below the control group (66.67% vs. 70.83%), and the ERD/ERS profile helps explain this result. The Music group shifted from Alpha ERD (observed in BL01-BL03) to pronounced Alpha ERS, indicating a transition toward reduced concentration or potential distraction. At the same time, Theta ERS in the Music group reflects increased working memory engagement, a combination that may be counterproductive for a creative, generative task requiring flexible ideation rather than focused error detection. It should be noted that BL04 was consistently the last block performed by all participants, so cumulative fatigue represents a plausible alternative or contributing explanation for the performance decline, independent of the neural interpretation offered above. Disentangling task-order fatigue effects from task-type effects was beyond the scope of this study and is an important direction for future work employing counterbalanced task ordering.

These findings suggest that the regulatory effect of music has an important boundary condition: when tasks require open-ended generation and verbal-creative reasoning, the same auditory stimulus that stabilizes attention in analytical contexts may instead compete for the cognitive resources necessary for free-form thinking. This interpretation aligns with Furnham and Bradley's [14] observation that background stimuli can differentially affect performance depending on task demands, and extends this perspective to the context of software testing.

5.2 Integration with Prior EEG Research in Software Engineering

Our findings both replicate and extend prior EEG-based studies in software engineering. Calcagno et al. [8] reported elevated Delta, Theta, and Beta power during code comprehension, interpreting these as markers of working memory load and cognitive alertness. Our results show that music systematically lowers frontal Delta — with statistically significant, large effects — which, taken together with Calcagno et al.'s framework, implies that music reduces the cognitive expenditure required for the same comprehension tasks, consistent with the concept of neural efficiency. Lin et al. [29] and Peitek et al. [39] established that high-performing programmers exhibit lower cognitive load signatures (reduced Beta, increased Alpha) compared to novices. The Music group's neural profile shares characteristics with these high-performance signatures, suggesting that music may partially emulate, at the neural level, the efficiency gains typically associated with greater expertise.

5.3 Practical Implications

The findings of this study can be translated into concrete strategies aimed at enhancing both the work environment and the performance of software testers.

Task-contingent use of music: A single, generalized recommendation (e.g., "use music at work") is insufficient. Practitioners should consider adopting music selectively, during analytical activities such as debugging, code review, or fault localization, while pausing it for open-ended tasks such as test design or exploratory testing, where it may induce attentional drift.

Music as a low-cost cognitive intervention: The consistently large Delta effect sizes observed across nearly all tasks, combined with the behavioral improvements reported in Section 4.2, support positioning music as a feasible, low-cost, and non-invasive cognitive aid, particularly for high-volume, repetitive testing activities where sustained attention is critical.

Scheduling and cognitive load management: The gradient of task difficulty observed across blocks, with music providing the most differentiated neural support in the most demanding task (BL03), suggests it may be especially beneficial during cognitively taxing testing phases, potentially buffering against fatigue-related performance decline.

5.4 Implications for Academia

This pioneering study opens a novel line of inquiry at the intersection of neuroscience, software engineering, and music.

Establishing a New Research Line: As the first EEG study on software testers, this work provides a methodological foundation for future research, including a validated set of 12 testing tasks and publicly available artifacts and datasets (see Section Artifacts Availability).

Exploring Individual Differences: While the musical stimulus was standardized, individual factors such as music preference and personality may mediate its effects [14] and warrant investigation in future work.

Adopting Multimodal Approaches: Future research should extend this multimodal design by integrating EEG with complementary techniques such as eye tracking [39], and by including more diverse, realistic testing activities—such as exploratory testing, security testing, or test suite refactoring.

6 Threats to Validity

In this section, we discuss potential threats to the validity of our study, following the guidelines proposed by Wohlin et al. [64].

Internal Validity: The controlled experimental setting may not fully replicate the complexity and dynamic nature of real-world software testing, where problem-solving demands and contextual variability are considerably higher [1]. Moreover, the use of short, simplified code snippets, necessary to accommodate EEG recording constraints, may not have consistently elicited the cognitive load levels found in real tasks. However, internal validity was strengthened through a controlled protocol: participants were randomly assigned to music or non-music groups, performed identical tasks, and followed standardized procedures. The use of a fixed instrumental playlist also helped reduce variability due to individual music preferences, improving the reliability of the observed effects. A further threat concerns the potential influence of the EEG apparatus itself on participant behavior: being monitored via a 64-channel cap may have induced heightened self-awareness or altered natural task engagement (a Hawthorne-like effect [32]), independent of the music condition. Since this effect would presumably apply similarly to both groups, it is unlikely to explain the between-group differences observed, but it may affect the generalizability of absolute performance and neural values to unmonitored, real-world settings.

External Validity: Generalizability is constrained by the modest sample size and unbalanced gender distribution, both influenced by the requirement for participants with software testing expertise and the logistical demands of EEG-based studies. Although the sample size constrains statistical inference, the primary goal of this study was to assess the feasibility of applying neuroscience methods—particularly EEG—in studies involving professional software testers. Future studies should aim for larger samples of industry professionals to enable deeper subgroup analysis. Still, recruiting experienced professionals from industry settings supports the practical relevance of our findings and supports their potential applicability in real-world software testing contexts.

Construct Validity: A central threat to construct validity is the challenge of capturing the complexity of real testing activities in a controlled experimental environment. Although we used a commonly adopted programming language and tasks aligned with industry practice, experimental tasks inevitably simplify real-world conditions. To address this, we conducted iterative pilot studies to fine-tune task difficulty, ensuring they were neither too simple nor unrealistically complex. This calibration, combined with established methodologies from software engineering and neuroscience, supported valid assessments of cognitive load and performance. Additionally, this study did not collect subjective measures of workload, stress, enjoyment, or perceived concentration (e.g., via instruments such as the NASA-TLX), which could help explain or corroborate the observed neural and behavioral patterns. Future work should incorporate such self-report measures to complement the physiological and performance data.

7 Conclusion

This study presents the first EEG-based investigation of how instrumental background music influences the neural activity and task performance of professional software testers. By combining power spectral density, ERD/ERS analysis, and behavioral accuracy across four testing tasks, we provide neurophysiologically grounded evidence that music functions as a task-contingent cognitive modulator.

The most robust neurophysiological finding concerns the Delta band, which showed significantly lower frontal power in the Music group across all task blocks ($d = -1.28$ to -1.69), indicating sustained neural readiness throughout the experiment. Theta reached significance in BL02 and, more tentatively, in BL03 (see Section 5.1 for a discussion of the discrepancy between the Mann-Whitney and t-test results in this block), while Alpha and Beta showed large directional effect sizes ($d \approx -0.81$ to -0.88) consistent with a flow-like attentional state, though underpowered for formal confirmation. Behaviorally, the Music group outperformed controls in comprehension (+8.33 pp), syntax (+12.5 pp), and logic error detection (+12.5 pp). The exception — test case creation (BL04, −4.16 pp) — is not an anomaly but a theoretically meaningful boundary condition: Alpha ERS in that block is directionally consistent with attentional drift precisely when open-ended generative reasoning was required, suggesting that music's benefits may be specific to convergent analytical tasks, a hypothesis that future work with larger samples should formally test.

These findings extend prior EEG research in software engineering to a previously unstudied professional role and provide a replicable protocol, a validated task set of 12 testing activities, and publicly available artifacts for future research. Key limitations include the small sample ($n = 14$), fixed playlist, and simplified code snippets, which constrain generalizability. Future work should pursue larger-sample replications to confirm Alpha and Beta effects, experimentally isolate the music-by-task-type interaction observed at BL04, and examine individual factors such as music preference and personality. As cognitive demands on software testers continue to grow, understanding how environmental factors modulate tester cognition carries direct practical relevance for software quality.

Artifact Availability

For replication purposes, all source code used for analysis and experimental tasks is publicly available, the supplementary materials and source code can be accessed at: https://github.com/DhyegoTavares/eeg-music-testing-replication. The raw and anonymized EEG data is available on Zenodo DOI: https://doi.org/10.5281/zenodo.22119099.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. It was also supported in part by FAPESB INCITE PIE0002/2022 grant, and INES.IA (National Institute of Science and Technology for Software Engineering Based on and for Artificial Intelligence) www.ines.org.br, CNPq grant 408817/2024-0.

References

[1] Pavlo Antonenko, Fred Paas, Roland Grabner, and Tamara Van Gog. 2010. Using electroencephalography to measure cognitive load. *Educational psychology review* 22 (2010), 425–438.

[2] Anum Asif, Muhammad Majid, and Syed Muhammad Anwar. 2019. Human stress classification using EEG signals in response to music tracks. *Computers in biology and medicine* 107 (2019), 182–196.

[3] Laura Barton, Gulipek Candan, Thomas Fritz, Thomas Zimmermann, and Gail C Murphy. 2019. The sound of software development: Music listening among software engineers. *IEEE Software* 37, 2 (2019), 78–85.

[4] Erol Başar, Canan Başar-Eroglu, Sirel Karakaş, and Martin Schürmann. 2001. Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International journal of psychophysiology* 39, 2-3 (2001), 241–248.

[5] Adnan Mehmood Bhatti, Muhammad Majid, Syed Muhammad Anwar, and Bilal Khan. 2016. Human emotion recognition and analysis in response to audio music using brain signals. *Computers in Human Behavior* 65 (2016), 267–275.

[6] Neil C. C. Brown, Pierre Weill-Tessier, Maksymilian Sekula, Alexandra-Lucia Costache, and Michael Kölling. 2022. Novice Use of the Java Programming Language. *ACM Trans. Comput. Educ.* 23, 1, Article 10 (dec 2022), 24 pages.

[7] Teresa Busjahn, Roman Bednarik, Andrew Begel, Martha Crosby, James H Paterson, Carsten Schulte, Bonita Sharif, and Sascha Tamm. 2015. Eye movements in code reading: Relaxing the linear order. In *2015 ieee 23rd international conference on program comprehension*. IEEE, Florence, Italy, 255–265.

[8] Alessandra Calcagno, Stefania Coelli, Ricardo Couceiro, João Durães, Caterina Amendola, Ileana Pirovano, Rebecca Re, and Anna Maria Bianchi. 2020. EEG monitoring during software development. In *2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON)*. IEEE, 325–329.

[9] Camila Cosmo, Amin Zandvakili, Nicholas J. Petrosino, Thaise Graziele L. de O. Toutain, José Garcia Vivas Miranda, and Noah S. Philip. 2023. Examining the neural mechanisms of rTMS: a naturalistic pilot study of acute and serial effects in pharmacoresistant depression.

[10] Igor Crk, Timothy Kluthe, and Andreas Stefik. 2015. Understanding programming expertise: an empirical study of phasic brain wave changes. *ACM Transactions on Computer-Human Interaction (TOCHI)* 23, 1 (2015), 1–29.

[11] Madeline Endres, Zachary Karas, Xiaosu Hu, Ioulia Kovelman, and Westley Weimer. 2021. Relating Reading, Visualization, and Coding for New Programmers: A Neuroimaging Study. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. IEEE, 600–612.

[12] Janet Feigenspan and Norbert Siegmund. 2012. Supporting comprehension experiments with human subjects. In *2012 20th IEEE International Conference on Program Comprehension (ICPC)*. IEEE, 244–246.

[13] Davide Fucci, Daniela Girardi, Nicole Novielli, Luigi Quaranta, and Filippo Lanubile. 2019. A replication study on code comprehension and expertise using lightweight biometric sensors. In *2019 IEEE/ACM 27th International Conference on Program Comprehension (ICPC)*. IEEE, 311–322.

[14] Adrian Furnham and Anna Bradley. 1997. Music while you work: The differential distraction of background music on the cognitive test performance of introverts and extraverts. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 11, 5 (1997), 445–455.

[15] Diana Guadalupe González-Rodríguez, Dulce Martinez-Peon, Xochitl Angélica Ortiz Jiménez, Juan Fernando Góngora-Rivera, Francisco Gerardo Benavides-Bravo, Brayan Soria Rodríguez, and Mariana Ruíz Velázquez. 2022. EEG-Based Drivers Mental Fatigue Detection Using ERD/ERS and Hurst Exponent. In *Proceedings of the 6th International Conference on Medical and Health Informatics* (Virtual Event, Japan) *(ICMHI '22)*. Association for Computing Machinery, New York, NY, USA, 159–162.

[16] William S Greenwell, Elisabeth A Strunk, and John C Knight. 2004. Failure analysis and the safety-case lifecycle. In *Human Error, Safety and Systems Development: IFIP 18th World Computer Congress TC13/WC13. 5 7th Working Conference on Human Error, Safety and Systems Development 22–27 August 2004 Toulouse, France.* Springer, 163–176.

[17] Gaurav Gupta, Sérgio Pequito, and Paul Bogdan. 2018. Re-thinking EEG-based non-invasive brain interfaces: Modeling and analysis. In *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, 275–286.

[18] Gao Hao, Haytham Hijazi, João Durães, Júlio Medeiros, Ricardo Couceiro, Chan Tong Lam, César Teixeira, João Castelhano, Miguel Castelo Branco, Paulo Carvalho, et al. 2023. On the accuracy of code complexity metrics: A neuroscience-based guideline for improvement. *Frontiers in Neuroscience* 16 (2023), 1065366.

[19] Jiancheng Hou, Bei Song, Andrew CN Chen, Changan Sun, Jiaxian Zhou, Haidong Zhu, and Theodore P Beauchaine. 2017. Review on neural correlates of emotion regulation and music: Implications for emotion dysregulation. *Frontiers in Psychology* 8 (2017), 501.

[20] Yoshiharu Ikutani and Hidetake Uwano. 2014. Brain activity measurement during program comprehension with NIRS. In *15th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. IEEE, Las Vegas, NV, USA, 1–6.

[21] Rozita Kadar, Syarifah Adilah Mohamed Yusoff, Saiful Nizam Warris, and Mohd Saifulnizam Abu Bakar. 2021. Students Assessments in Learning Programming based on Bloom's Taxonomy. *Journal of Computing Research and Innovation* 6, 3 (2021), 13–21.

[22] Wolfgang Klimesch. 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews* 29, 2-3 (1999), 169–195.

[23] Wolfgang Klimesch. 2012. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences* 16, 12 (2012), 606–617.

[24] Stefan Koelsch. 2014. Brain correlates of music-evoked emotions. *Nature reviews neuroscience* 15, 3 (2014), 170–180.

[25] Ryan Krueger, Yu Huang, Xinyu Liu, Tyler Santander, Westley Weimer, and Kevin Leach. 2020. Neurological divide: an fMRI study of prose and code writing. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*. 678–690.

[26] SeolHwa Lee, Andrew Matteson, Danial Hooshyar, SongHyun Kim, JaeBum Jung, GiChun Nam, and HeuiSeok Lim. 2016. Comparing programming language comprehension between novice and expert programmers using eeg analysis. In *2016 IEEE 16th international conference on bioinformatics and bioengineering (BIBE)*. IEEE, Taichung, Taiwan, 350–355.

[27] Teresa Lesiuk. 2000. The effect of music listening on a computer programming task. *Journal of Computer Information Systems* 40, 3 (2000), 50–57.

[28] Teresa Lesiuk. 2005. The effect of music listening on work performance. *Psychology of music* 33, 2 (2005), 173–191.

[29] Yu-Tzu Lin, Yi-Zhi Liao, Xiao Hu, and Cheng-Chih Wu. 2021. Eeg activities during program comprehension: An exploration of cognition. *IEEE Access* 9 (2021), 120407–120421.

[30] Henry B Mann and Donald R Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics* (1947), 50–60.

[31] Christoph Mayr-Dorn, Michael Vierhauser, Stefan Bichler, Felix Keplinger, Jane Cleland-Huang, Alexander Egyed, and Thomas Mehofer. 2021. Supporting Quality Assurance with Automated Process-Centric Quality Constraints Checking. In *Proceedings of the 43rd International Conference on Software Engineering (ICSE '21)*. IEEE Press, Madrid, Spain, 1298–1310.

[32] Rob McCarney, James Warner, Steve Iliffe, Robbert Van Haselen, Mark Griffin, and Peter Fisher. 2007. The Hawthorne Effect: a randomised, controlled trial. *BMC medical research methodology* 7, 1 (2007), 30.

[33] Júlio Medeiros, Ricardo Couceiro, Gonçalo Duarte, João Durães, João Castelhano, Catarina Duarte, Miguel Castelo-Branco, Henrique Madeira, Paulo de Carvalho, and César Teixeira. 2021. Can EEG be adopted as a neuroscience reference for assessing software programmers' cognitive load? *Sensors* 21, 7 (2021), 2338.

[34] Vojkan Mihajlović, Bernard Grundlehner, Ruud Vullers, and Julien Penders. 2014. Wearable, wireless EEG solutions in daily life applications: what are we missing? *IEEE journal of biomedical and health informatics* 19, 1 (2014), 6–21.

[35] Kimberly Sena Moore. 2013. A systematic review on the neural effects of music on emotion regulation: Implications for music therapy practice. *Journal of music therapy* 50, 3 (2013), 198–242.

[36] Makayla Moster, Aarav Chandra, Christal Chu, Weiyi Liu, and Paige Rodeghero. 2022. In the Zone: An Analysis of the Music Practices of Remote Software Developers. In *Proceedings of the 16th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement* (Helsinki, Finland) *(ESEM '22)*. Association for Computing Machinery, New York, NY, USA, 313–318.

[37] Greg R Oldham, Anne Cummings, Leann J Mischel, James M Schmidtke, and Jing Zhou. 1995. Listen while you work? Quasi-experimental relations between personal-stereo headset use and employee work responses. *Journal of Applied Psychology* 80, 5 (1995), 547.

[38] Norman Peitek, Sven Apel, Chris Parnin, André Brechmann, and Janet Siegmund. 2021. Program Comprehension and Code Complexity Metrics: An fMRI Study. In *Proceedings of the 43rd International Conference on Software Engineering* (Madrid, Spain) *(ICSE '21)*. IEEE Press, 524–536.

[39] Norman Peitek, Annabelle Bergum, Maurice Rekrut, Jonas Mucke, Matthias Nadig, Chris Parnin, Janet Siegmund, and Sven Apel. 2022. Correlates of programmer efficacy and their link to experience: a combined EEG and eye-tracking study. In *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (Singapore, Singapore) *(ESEC/FSE 2022)*. Association for Computing Machinery, New York, NY, USA, 120–131.

[40] Norman Peitek, Janet Siegmund, Sven Apel, Christian Kästner, Chris Parnin, Anja Bethmann, Thomas Leich, Gunter Saake, and André Brechmann. 2018. A look into programmers' heads. *IEEE Transactions on Software Engineering* 46, 4 (2018), 442–462.

[41] Norman Peitek, Janet Siegmund, Chris Parnin, Sven Apel, Johannes C. Hofmeister, and André Brechmann. 2018. Simultaneous measurement of program comprehension with fMRI and eye tracking: a case study. In *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* (Oulu, Finland) *(ESEM '18)*. Association for Computing Machinery, New York, NY, USA, Article 24, 10 pages.

[42] Gert Pfurtscheller and A Aranibar. 1977. Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalography and clinical neurophysiology* 42, 6 (1977), 817–826.
[43] Gert Pfurtscheller and FH Lopes Da Silva. 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology* 110, 11 (1999), 1842–1857.
[44] R. Pressman. 2014. *Software Engineering: A Practitioner's Approach*. McGraw-Hill Education. https://books.google.com.br/books?id=dXlzCgAAQBAJ
[45] Nornadiah Mohd Razali and Yap Bee Wah. 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics* 2, 1 (2011), 21–33.
[46] Patrick Rempel, Patrick Mäder, Tobias Kuschke, and Jane Cleland-Huang. 2014. Mind the gap: assessing the conformance of software traceability to relevant guidelines. In *Proceedings of the 36th International Conference on Software Engineering* (Hyderabad, India) *(ICSE 2014)*. Association for Computing Machinery, New York, NY, USA, 943–954.
[47] Saeid Sanei and J. A. Chambers. 2009. *EEG signal processing*. John Wiley and Sons.
[48] Paul Sauseng, Wolfgang Klimesch, Manuel Schabus, and Michael Doppelmayr. 2005. Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International journal of Psychophysiology* 57, 2 (2005), 97–103.
[49] Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3-4 (1965), 591–611.
[50] Janet Siegmund, André Brechmann, Sven Apel, Christian Kästner, Jörg Liebig, Thomas Leich, and Gunter Saake. 2012. Toward measuring program comprehension with functional magnetic resonance imaging. In *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering* (Cary, North Carolina) *(FSE '12)*. Association for Computing Machinery, New York, NY, USA, Article 24, 4 pages.
[51] Janet Siegmund, Christian Kästner, Sven Apel, Chris Parnin, Anja Bethmann, Thomas Leich, Gunter Saake, and André Brechmann. 2014. Understanding understanding source code with functional magnetic resonance imaging. In *Proceedings of the 36th International Conference on Software Engineering* (Hyderabad, India) *(ICSE 2014)*. Association for Computing Machinery, New York, NY, USA, 378–389.
[52] Janet Siegmund, Christian Kästner, Jörg Liebig, Sven Apel, and Stefan Hanenberg. 2014. Measuring and modeling programming experience. *Empirical Software Engineering* 19, 5 (2014), 1299–1334.
[53] Janet Siegmund, Christian Kästner, Sven Apel, André Brechmann, and Gunter Saake. 2013. Experience from measuring program comprehension - toward a general framework.
[54] Janet Siegmund, Norman Peitek, Chris Parnin, Sven Apel, Johannes Hofmeister, Christian Kästner, Andrew Begel, Anja Bethmann, and André Brechmann. 2017. Measuring neural efficiency of program comprehension. In *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering* (Paderborn, Germany) *(ESEC/FSE 2017)*. Association for Computing Machinery, New York, NY, USA, 140–150.
[55] Aleksi J Sihvonen and Teppo Särkämö. 2022. Music processing and amusia. *Handbook of Clinical Neurology* 187 (2022), 55–67.
[56] Roselane Silva Farias, Iftekhar Ahmed, and Eduardo Santana de Almeida. 2025. What Makes a Great Software Quality Assurance Engineer? *IEEE Transactions on Software Engineering* (2025), 1–20.
[57] WAS Smith. 1961. Effects of industrial music in a work situation requiring complex mental activity. *Psychological Reports* 8, 1 (1961), 159–162.
[58] Student. 1908. The probable error of a mean. *Biometrika* (1908), 1–25.
[59] Eoin Thomas, Joan Fruitet, and Maureen Clerc. 2012. Investigating brief motor imagery for an ERD/ERS based BCI. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2929–2932.
[60] Fuyu Ueno and Sotaro Shimada. 2023. Inter-subject correlations of EEG reflect subjective arousal and acoustic features of music. *Frontiers in Human Neuroscience* 17 (2023).
[61] Barbara Weber, Thomas Fischer, and René Riedl. 2021. Brain and autonomic nervous system activity measurement in software engineering: A systematic literature review. *Journal of Systems and Software* 178 (2021), 110946.
[62] Peter Welch. 2003. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics* 15, 2 (2003), 70–73.
[63] Emma Winston and Laurence Saywood. 2019. Beats to relax/study to: Contradiction and paradox in lofi hip hop. *IASPM Journal* 9, 2 (2019), 40–54.
[64] Claes Wohlin, Per Runeson, Martin Höst, Magnus C Ohlsson, Björn Regnell, and Anders Wesslén. 2012. *Experimentation in software engineering*. Springer Science & Business Media.
[65] Marvin Wyrich, Justus Bogner, and Stefan Wagner. 2023. 40 years of designing code comprehension experiments: A systematic mapping study. *Comput. Surveys* 56, 4 (2023), 1–42.
[66] Martin K-C Yeh, Dan Gopstein, Yu Yan, and Yanyan Zhuang. 2017. Detecting and comparing brain activity in short program comprehension using EEG. In *2017 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–5.