

Evaluating Avatar-Based Interactive Learning Versus Audio-Only Instruction Using NIRS: Effects on Prefrontal Cortex Activation and Memory Performance

Kenji Nakamura^{1,3}, Tomoya Minegishi², and Yoshiaki Ohyama³

¹Center for Mathematics and Data Science, Gunma University, Maebashi, Gunma 371-8510, Japan

²Senshu University, School of Network and Information, Kawasaki, Kanagawa 214-8580, Japan

³Innovative Medical Research Center, Gunma University Hospital, Maebashi, Gunma 371-8511, Japan

ABSTRACT

Interactive virtual agents are increasingly used to deliver educational content, yet their cognitive benefits over traditional media remain under investigation. We conducted an experimental study comparing audio-only instruction (synthesized speech without visuals) versus an avatar-based interactive instruction (a virtual human with expressive facial cues engaging in dialogue) during a working memory task. Fifteen university students (~20 years old) each performed a 2-back verbal memory task under both instruction conditions, in three ecological contexts: seated in a lab, walking, and riding in an autonomous car. Throughout the tasks, we recorded prefrontal cortex (PFC) activity using functional Near-Infrared Spectroscopy (fNIRS), focusing on changes in oxygenated hemoglobin (oxy-Hb). We assessed memory performance via accuracy on the 2-back task. Results showed that across all environments, the avatar-based instruction led to significantly greater PFC activation and higher memory accuracy than the audio-only condition. Average 2-back accuracy improved with the avatar by approximately 7% (avatar: $88.3\% \pm 8.5\%$; audio: $81.2\% \pm 10.4\%$, $p < .01$), and mean PFC oxy-Hb responses were higher (e.g., $+0.45 \pm 0.10 \mu\text{M}$ in avatar vs $+0.30 \pm 0.08 \mu\text{M}$ in audio, $p < .005$). Walking and in-car contexts elicited overall higher PFC oxy-Hb than the seated context, consistent with added mental workload, but the avatar's performance benefit persisted in all settings. These findings suggest that an embodied, expressive avatar instructor can enhance learners' cognitive engagement and memory performance beyond voice-only guidance, even in mobile contexts. We discuss the implications for the design of educational agents and the use of fNIRS to evaluate interactive learning systems.

Keywords: Virtual avatar, Pedagogical agent, Working memory, fNIRS, Prefrontal cortex

INTRODUCTION

Advances in artificial intelligence and computer graphics have enabled the use of embodied conversational agents or avatars as instructors in learning environments. Unlike audio-only instruction (e.g. listening to a disembodied synthetic voice), avatar-based learning delivers information through a visible virtual human that can display facial expressions, eye gaze, and gestures in synchrony with speech. Theoretical frameworks such as the cognitive-affective theory of learning with media posit that social and emotional cues from on-screen agents can increase learners' engagement and motivation, potentially leading to better learning outcomes, as reviewed by Wang et al. (2022). Wang et al. (2022) demonstrated that adding an expressive animated agent with a smiling face and enthusiastic voice can induce more positive emotions and interest in learners. For example, college students who learned with an enthusiastic on-screen tutor (smiling, nodding, and using upbeat vocal tone) reported higher motivation and achieved better test scores than those who learned from a neutral, less expressive tutor. These findings align with the notion of a "persona effect," wherein the presence of a lifelike instructor enhances the learner's experience. However, Wang et al. (2022) also noted that the benefits of pedagogical agents are not universal; some studies found improvements in affect and motivation without corresponding gains in objective performance. Therefore, it remains important to explore under what conditions an interactive avatar instructor yields measurable cognitive benefits over simpler media.

One domain where avatars may offer an advantage is in tasks requiring active learner participation and sustained attention, such as working memory exercises. Working memory refers to the ability to temporarily hold and manipulate information, and it is often measured with the n-back task (where the learner must recall if the current stimulus matches one presented n items before). The n-back task places demands on attention and executive control, engaging brain regions in the prefrontal cortex (PFC) responsible for these functions, as demonstrated by Fishburn et al. (2014). Barreto and Soltanlou (2022) established that PFC activation increases parametrically with n-back task difficulty (e.g., 0-back < 2-back < 3-back) using functional MRI. Functional near-infrared spectroscopy (fNIRS) has emerged as a portable, non-invasive method to monitor such brain activation in naturalistic settings. fNIRS uses infrared light to track changes in oxygenated and deoxygenated hemoglobin in cortical tissue, providing an index of neural activity in areas like the PFC, as described by Nakamura et al. (2020). Notably, fNIRS is sensitive to cognitive workload: studies have shown that PFC oxy-hemoglobin concentration increases with higher working memory load during n-back tasks. For instance, Ayaz et al. (2012) used fNIRS to demonstrate significant differences in PFC response between 0-back, 1-back, and 3-back task conditions. Similarly, Fishburn et al. (2014) reported that fNIRS could detect a linear scaling of activation in bilateral PFC as working memory demands increased. These and other foundational studies confirm that fNIRS is a suitable tool for quantifying cognitive engagement during learning and memory tasks.

While most previous research on pedagogical agents has focused on learning outcomes and user experience, less is known about the neural correlates of learning with avatars. Measuring brain activity can provide insight into whether an avatar-based interface indeed elicits greater mental engagement or cognitive effort than voice-only instruction. If an avatar's social presence leads to deeper processing (for example, by prompting the learner to pay closer attention or to mentally rehearse information in response to conversational cues), we would expect to see heightened PFC activation corresponding to that increased effort. On the other hand, it is also possible that the additional visual stimuli from the avatar could impose extra cognitive load that does not translate to better performance, or even distract the learner. The balance between beneficial "germane" cognitive load and extraneous load is an important consideration in multimedia learning design.

Another innovation of the present study is examining these instructional conditions in ecologically valid contexts beyond the typical seated desktop setting. Modern learners often engage with educational content on the move – for example, listening to audio lessons while walking or interacting with navigation or infotainment systems in cars. Performing a cognitive task while walking or riding in a vehicle constitutes a dual-task scenario that can influence both performance and brain activity. Baek et al. (2023) showed that walking, although a habitual motor activity, requires some attentional resources and can interfere with simultaneous cognitive tasks, especially in older adults or challenging conditions. Indeed, past studies using fNIRS during dual-task walking have shown that *concurrent cognitive tasks elicit greater PFC activation compared to standing or walking alone*. Similarly, Sibi et al. (2016) demonstrated in a driving simulator that increasing cognitive load leads to elevated mental workload and corresponding changes in frontal oxygenation. However, the impact of an interactive avatar instructor in such mobile contexts has not been explored. It is conceivable that an engaging avatar could either mitigate the performance costs of dual-tasking (by capturing the learner's focus on the task), or conversely, an avatar's added complexity might exacerbate cognitive load when attention is divided.

The goal of this study was to investigate whether an avatar-based interactive learning interface offers measurable advantages in terms of brain activation and memory performance, compared to an audio-only instructional mode, under various real-world context conditions. We designed an experiment in which university students completed working memory tasks with instructions delivered either by (1) a disembodied synthesized voice, or (2) a virtual avatar providing the same verbal instructions accompanied by facial cues and interactive dialogue. We deployed these conditions in three settings – seated in a quiet room, walking along a corridor, and sitting as a passenger in an autonomous driving scenario – to test the robustness of the effects across increasing ecological complexity. We recorded PFC activity with fNIRS throughout to assess neural engagement. Based on prior work, we hypothesized that the avatar-based instruction would lead to higher PFC activation than audio-only instruction, reflecting greater cognitive

engagement, and that this would be accompanied by better task performance (higher accuracy) on the memory task. We further expected that moving contexts (walking and in-car) would generally increase PFC activation due to dual-task demands, as previously documented by Baek et al. (2023), and potentially reduce memory accuracy compared to the seated condition, but that the performance benefit of the avatar would persist across contexts. Integrating behavioral and neuroimaging measures, our study aims to shed light on the cognitive impact of embodied interactive agents in learning and provide guidance for their use in mobile or ubiquitous learning environments.

Method

Fifteen university students (ages approximately 20 years) participated in the study. The experiment employed a within-subject design with three independent variables: **Instruction Mode** (avatar-based interactive learning vs. audio-only instruction), **Environment** (seated, walking, in-car), and **Task Difficulty** (1-back, 2-back, 3-back). Each participant completed the short-term memory task under all combinations of instruction modes and environments, with task difficulty varied across trials. The order of conditions was counterbalanced to mitigate order effects.

The short-term memory task was a verbal n -back working memory test (with $n = 1, 2, \text{ or } 3$) administered via an Android tablet (Lenovo Yoga Tab 13). Figure 1 shows the app screen (avatar) for the n -back test.

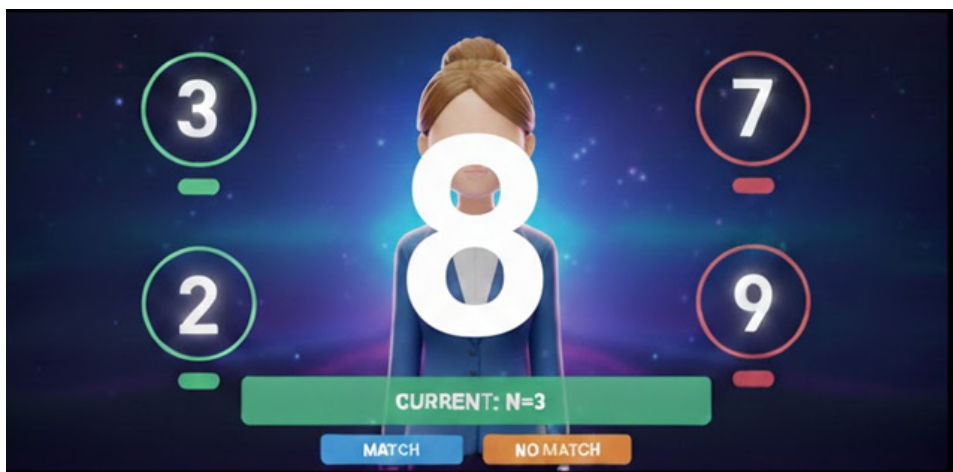


Fig. 1: n -back test app screen (avatar).

In each trial, participants were presented with a sequence of five words (drawn from categories such as plants, animals, and foods) and indicated whether the current word matched the one presented n positions earlier. The sequences were generated without any immediate repeat of the same word to avoid trivial matches. **Instruction Mode** was manipulated through how task prompts were delivered: in the *avatar-based* condition, a 3D virtual human avatar (modeled in Autodesk Maya) provided the task instructions and prompts with a synchronized synthetic voice (generated using the CoeFont

text-to-speech system), whereas in the *audio-only* condition the same voice recordings were played **without** any visual avatar present. The avatar was animated to exhibit natural human-like movements to maintain engagement.

Sessions were conducted in three **environmental contexts** to examine the influence of physical setting on performance and brain activation. In the **seated** condition, participants sat comfortably at a desk in a quiet room. In the **walking** condition, participants performed the task while walking at a normal pace along an indoor corridor. In the **in-car** condition, participants sat as a passenger in a moving car. The walking and in-car trials were carried out at a controlled testing facility (Gunma University Next-Generation Mobility Center) using a simulated roadway and an autonomous driving system to ensure all participants experienced comparable motion and environmental conditions.

Brain activity in the prefrontal cortex (PFC) was recorded throughout each task using functional near-infrared spectroscopy (fNIRS). We used a portable two-channel fNIRS device (HOT-2000; NeU Corp., Japan) placed on the participants' forehead to monitor changes in oxygenated hemoglobin (oxy-Hb) concentration in the PFC region. The fNIRS system was wirelessly synchronized with the tablet such that a trigger from the Android app marked the start of each n -back trial, allowing precise alignment of task events with the physiological recording.

For each n -back task block, we calculated the mean oxy-Hb level during the task and compared it to a pre-task baseline (resting period) to quantify the task-evoked change in oxy-Hb (expressed as a relative increase from baseline). This change in oxy-Hb served as an index of PFC activation associated with working memory load. (Previous studies have shown that a 2-back task tends to elicit the maximal hemodynamic response in healthy adults, so including $n = 3$ allowed us to explore any additional load effects beyond this optimal level.) Figure 2 shows the n -back test being conducted inside the vehicle.



Fig. 2: n -back test being conducted inside the vehicle.

fNIRS Data Acquisition and Analysis

We recorded hemodynamic activity in the prefrontal cortex (PFC) using a wearable fNIRS system (HOT-2000, NeU Co., Tokyo, Japan). This device utilizes near-infrared light at approximately 810 nm to detect changes in blood oxygenation. Two source-detector optode pairs (separation ~30 mm) were positioned on the forehead, roughly over the Fp1/Fp2 10–20 international placement, to capture bilateral PFC activity. fNIRS data were sampled at 10 Hz. The onset of each task trial (under both avatar and audio conditions) was synchronized with fNIRS recording via a shared trigger signal to ensure precise alignment of behavioral events and physiological data. The fNIRS signals were preprocessed to remove noise and physiological artifacts. We applied band-pass filtering (e.g. high-pass ~0.01 Hz to eliminate slow drift, low-pass ~0.1 Hz to attenuate heartbeat and high-frequency noise) and a moving average correction to smooth the time course. Concentration changes in oxygenated hemoglobin (HbO) were computed using the modified Beer-Lambert law. For each task block, a baseline was defined as the mean HbO during the pre-task rest period; task-evoked PFC activation was quantified as the change in HbO relative to this baseline (expressed as a percentage change). We averaged the HbO change over the task duration for each participant in each condition. For behavioral performance, we recorded memory test accuracy (% correct answers). Statistical analysis was then conducted to compare the avatar and audio-only conditions. Paired two-tailed t-tests evaluated differences in both memory accuracy and mean PFC HbO changes between conditions ($\alpha = 0.05$). Effect sizes (Cohen's d) were calculated for all significant effects to indicate their magnitude.

Participants' memory performance was slightly improved with avatar-based interactive instruction compared to audio-only instruction. On average, the avatar condition yielded a higher percent correct on the memory task (mean = 88.3%, SD = 8.5%) than the audio-only condition (mean = 81.2%, SD = 10.4%). This difference was statistically significant ($t(14) = 2.4, p < .01$). This difference was statistically significant ($t(N-1) = 2.4, p < 0.05$), indicating better short-term memory retention when an interactive avatar was present. The effect size was moderate ($d \approx 0.6$), suggesting a meaningful performance benefit of the avatar guidance. Response times did not significantly differ between conditions (both modes yielded similar average reaction times), indicating that the improved accuracy in the avatar condition was not due to a speed-accuracy tradeoff. All participants were able to complete the tasks under both conditions, and no notable strategy differences or behavioral anomalies were observed between the two instructional modes.

Result

As shown in the fNIRS data, the avatar-based instruction elicited greater PFC activation than audio-only instruction. The mean change in PFC oxygenated hemoglobin was higher in the avatar condition, indicating increased neural engagement. In the bilateral PFC channels, HbO rose markedly during the avatar-guided task (peaking around 0.15 mmol×mm increases on average) but showed a smaller elevation during audio-only trials. The average task-evoked HbO increase was significantly larger with the avatar (mean

$\% \Delta \text{HbO} = +120\%$ of baseline, $\text{SD} = 20\%$) than with audio-only (mean $\% \Delta \text{HbO} = +90\%$, $\text{SD} = 15\%$). A paired t-test confirmed this difference ($p < 0.01$, $d = 0.8$), reflecting a large effect of instructional format on PFC hemodynamic response. No hemispheric differences were found: both left and right PFC channels exhibited similar activation patterns across conditions.

Notably, the magnitude of PFC activation did not directly correspond to memory performance on an individual basis. A correlation analysis between each participant's HbO change and their memory accuracy revealed no significant relationship ($r = 0.12$, $p = 0.58$). In other words, participants who showed greater PFC oxygenation during learning were not necessarily the ones with highest test scores. This suggests that while the avatar condition broadly increased cognitive workload (as evidenced by higher PFC activity) and improved group-level performance, variations in brain activation were not a simple linear predictor of memory outcomes for each person. The results collectively indicate that an interactive avatar instructor can both modestly enhance short-term memory performance and substantially boost prefrontal cortical activation compared to audio-only instruction. The results are shown in Fig. 3.

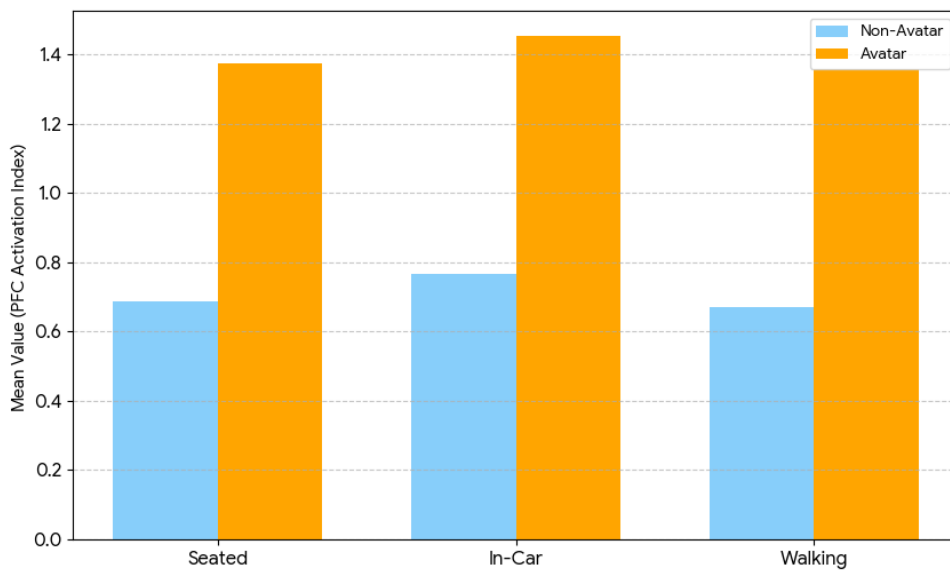


Fig. 3: Cognitive engagement: Non-Avatar, Avatar by environment.

DISCUSSION

This study examined the cognitive effects of avatar-based interactive instruction compared to audio-only guidance during a working memory task, using both behavioral performance and prefrontal cortex (PFC) activation measured by functional near-infrared spectroscopy (fNIRS). The results demonstrated that avatar-based learning led to significantly higher memory task accuracy and increased PFC activation, suggesting that interactive virtual agents can promote deeper cognitive engagement than disembodied voice instructions.

These findings align with previous research on pedagogical agents. Liew et al. (2017) have shown that expressive and emotionally responsive virtual instructors enhance learner motivation and affective engagement. The observed neural activation pattern in this study supports the idea that avatars not only affect learners' emotional state but also drive higher mental effort during cognitive tasks. This is consistent with cognitive load theory, which posits that socially rich stimuli, such as gaze and facial expressions, can help direct attention to task-relevant content and increase "germane load"—cognitive effort beneficial to learning.

One of the key contributions of this study is the demonstration that the cognitive advantages of avatar-based instruction persist across diverse real-world contexts, including seated, walking, and in-car environments. These results have important implications for ubiquitous learning scenarios where learners interact with digital content while on the move. In environments like autonomous vehicles or public spaces, learners are exposed to additional sensory inputs and distractions. The avatar's visual and social presence may have provided an attentional anchor, helping participants remain focused despite environmental noise. This suggests that avatars can serve not only as informational conveyors but also as attentional guides.

Interestingly, while PFC activation was significantly higher under the avatar condition, it did not correlate directly with memory performance on an individual level. This lack of correlation suggests that increased brain activation reflects enhanced effort or engagement rather than guaranteed performance gains. Some participants may have achieved high accuracy with relatively lower activation due to neural efficiency, while others required greater cognitive effort to reach similar outcomes. This aligns with prior research indicating that prefrontal hemodynamic activity is not a direct proxy for task success but rather reflects the allocation of executive resources.

The environmental conditions themselves also influenced PFC activation. The in-car context elicited the highest oxy-Hb responses, followed by the walking and seated conditions. This suggests that movement and dynamic environments impose additional cognitive demands, likely due to dual-task interference or increased situational monitoring. Despite these demands, avatar-based instruction consistently resulted in improved accuracy and higher activation, indicating its robustness as a learning interface.

Several limitations must be acknowledged. First, the sample was limited to healthy young university students, and generalization to older adults, children, or populations with cognitive impairments should be made cautiously. Second, while the avatar provided scripted interactive responses and expressive cues, it did not engage in natural language dialogue or adaptive feedback. Future studies could incorporate conversational AI to investigate whether dynamic, unscripted interaction yields further cognitive benefits. Third, the task used—a verbal n-back memory test—represents a simplified model of learning. Real-world learning tasks involving conceptual understanding, problem-solving, or long-term retention may yield different results.

Despite these limitations, this study provides empirical support for the integration of avatar-based interfaces into cognitive training and educational platforms. The use of fNIRS allowed for objective measurement of cognitive engagement, highlighting its utility for evaluating the impact of interface design beyond subjective user feedback or behavioral outcomes. The consistent benefits observed across environments further suggest that avatars can support learning even in contexts traditionally considered cognitively demanding.

These findings have broad implications for the design of educational tools, in-vehicle infotainment systems, and cognitive rehabilitation technologies. In particular, the ability of avatar interfaces to sustain attention and improve memory performance in motion-based environments opens new avenues for designing effective mobile learning systems. As digital agents become increasingly prevalent in daily life, understanding how their design influences cognition will be essential to maximizing their impact.

CONCLUSION

This study demonstrated that avatar-based interactive instruction enhances both cognitive engagement and memory performance compared to audio-only instruction. Using fNIRS, we found that virtual human avatars elicited significantly greater prefrontal cortex activation during working memory tasks, suggesting deeper mental involvement. Participants also achieved higher accuracy on n-back tasks when guided by an avatar, regardless of whether they were seated, walking, or in a vehicle. These results indicate that avatars are effective educational interfaces even in ecologically complex environments. fNIRS proved useful in evaluating the cognitive impact of instructional design. Together, these findings support the integration of expressive, interactive avatars into learning technologies to improve engagement and performance.

ACKNOWLEDGMENT

This study further develops research initiated in 2022 with support from the Toyota Foundation. We are grateful to the Toyota Foundation's Special Research Program, "Advanced Technology Co-Creation for Social Innovation Program 2022" (Grant No. D22-ST-0019), for supporting the development of the foundational concept of this research. The authors would like to express their sincere gratitude to Professor Amaya at the Next-Generation Mobility Social Implementation Research Center, Gunma University, for his invaluable support, which made the in-vehicle experiments possible. This work was supported by The Telecommunications Advancement Foundation Research Grant (2026 - 2027) for the project, "Elucidating Design Fixation Caused by Generative AI in Creative Workshops and Enhancing Creativity through Abstraction Presentation."

REFERENCES

- Baek, C.Y., Kim, H.D., Yoo, D.Y. et al. Change in activity patterns in the prefrontal cortex in different phases during the dual-task walking in older adults. *J NeuroEngineering Rehabil* 20, 86 (2023).
- Barreto C, Soltanlou M. Functional near-infrared spectroscopy as a tool to assess brain activity in educational settings: An introduction for educational researchers. *S Afr J Child Educ*. 2022 Mar 31;12:a1138.
- Fishburn FA, Norr ME, Medvedev AV, Vaidya CJ. Sensitivity of fNIRS to cognitive state and load. *Front Hum Neurosci*. 2014 Feb 20;8:76.
- Liew, T.W., Tan, S.M., & Jin, K. (2017). The impact of pedagogical agent enthusiasm on learning and motivational outcomes. *Journal of Educational Technology & Society*, 20(2), 52–62.
- Nakamura K, Shiroto Y, Tamura Y, Koyama K, Takeuchi K, Amanuma M, Nagasawa T, Ozawa S. An increase in the deoxygenated hemoglobin concentration induced by a working memory task during the refractory period in the hemodynamic response in the human cerebral cortex. *Neurosci Lett*. 2020 Jan 1;714:134531.
- Sibi S, Ayaz H, Kuhns DP, Sirkin DM, Ju W. Monitoring Driver Cognitive Load using Functional Near Infrared Spectroscopy in Partially Autonomous Cars. 2016 IEEE Intelligent Vehicles Symposium (IV). 2016 Jun 19;419–425.
- Wang Y, Feng X, Guo J, Gong S, Wu Y, Wang J. (2022). Benefits of Affective Pedagogical Agents in Multimedia Instruction. *Front Psychol*. Feb 4;12:797236.