Visual Cues in Augmented Reality Piano Playing

Rebecca Yee
Carleton University
Ottawa, Canada
RebeccaYee@cmail.carleton.ca

Christopher Du
Carleton University
Ottawa, Canada
ChristopherDu@cmail.carleton.ca

Robert J. Teather
Monash University
Melbourne, Australia
Rob.Teather@monash.edu

Hamza Bashandy Taalab
Carleton University
Ottawa, Canada
HamzaBashandyTaalab@cunet.carleton.ca

Anthony Scavarelli
Toronto Metropolitan University
Toronto, Canada
anthony.scavarelli@torontomu.ca

Abstract

We present an augmented reality (AR) piano-playing system investigating the effectiveness of three visual cue strategies for guided key-press interaction. Using a within-subjects design, we evaluated reaction time, error rate, mental workload, and system usability. Although key-pressing performance yielded no statistically significant differences, descriptive trends and qualitative feedback indicate that complex multi-target visual cues impose higher cognitive demand. Spatial alignment limitations further impacted accuracy, highlighting the necessity of high-precision tracking and contextual visual aids for AR instrument learning.

CCS Concepts

• **Human-centered computing → Empirical studies in visualization**; **Mixed / augmented reality**.

Keywords

Augmented Reality, Piano, Learning, Visual Cues

ACM Reference Format:
Rebecca Yee, Christopher Du, Robert J. Teather, Hamza Bashandy Taalab, and Anthony Scavarelli. 2026. Visual Cues in Augmented Reality Piano Playing. In *ACM Symposium on Spatial User Interaction (SUI '26), October 10–11, 2026, Bari, Italy.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3822518.3839316

1 Introduction

While augmented reality (AR) applications for learning musical instruments are increasingly common [2], complex visual cues can easily overwhelm users. To evaluate the trade-offs between predictive visual benefits and cognitive load, we compared three AR cueing techniques for guided piano key-pressing (Figure 1): **SINGLE** (highlights the current target green), **DIRECTION** (adds a 3D arrow pointing to the next target), and **STOPLIGHT** (highlights three upcoming keys in a green-yellow-red sequence).

This work is licensed under a Creative Commons Attribution 4.0 International License.
SUI '26, Bari, Italy
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2812-9/2026/10
https://doi.org/10.1145/3822518.3839316

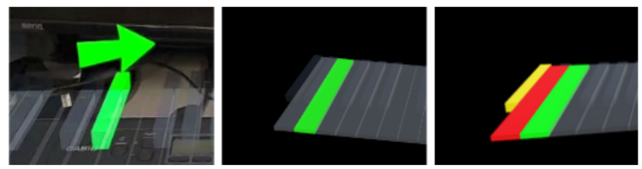

Figure 1: Visual cue techniques employed in this study. From left to right: DIRECTION, SINGLE, and STOPLIGHT.

To determine whether added visual complexity benefits AR-guided playing, we conducted a within-subjects experiment measuring reaction time (RT), error rate, mental workload, and system usability.

2 Related Work

AR/VR integration has attracted growing interest in music education [1]. Common methods include visually priming the piano keys to be pressed with virtual cues that "fly into" the physical keys to indicate when to press them [6] and AR augmentation to highlight the target keys on the physical keyboard or piano [3]. Schwandt et al. [4] found that visual cues in a VR space-mission simulation improved task performance and reduced mental workload, with visual-only cues outperforming audiovisual ones. Similarly, Zhang et al. [8] reported that visual cues substantially reduced task time and mental workload in a book-search task. Volmer et al. [7] employed EEG and demonstrated that predictive visual cues resulted in the lowest mental workload. In AR and MR environments, visual cues play a facilitative role by reducing task cognitive demands by offloading processes such as visual search, vigilance, and situational awareness onto them.

3 User Study

3.1 Participants, Apparatus, and Task

We recruited 9 university participants (6M, 3F; aged 19–31, $M = 24.5$, $SD = 3.36$). Eight had AR/VR experience; five had piano experience (1–16 years). The AR app was developed in Unity 6 using the Meta XR SDK and MidiJack, running on a Quest 3 connected to a Casiotone CT-S300 MIDI keyboard. To compensate for tracking drift, semi-transparent virtual overlays hovered 5 cm above the physical keys [5]. Random target keys spanned the keyboard. **SINGLE** highlighted one target; **DIRECTION** added a 3D arrow pointing to the next key; **STOPLIGHT** highlighted three future targets in a

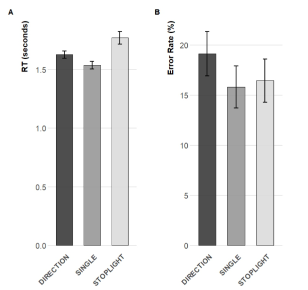

Figure 2: (A) RT in seconds and (B) error rate as a percent. Lower is better.

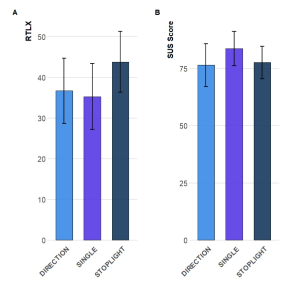

Figure 3: (A) RTLX scores and (B) SUS scores for the three visual cue types.

green-yellow-red sequence. Correct presses instantly triggered the next target; errors flashed grey.

3.2 Design and Procedure

We used a counterbalanced within-subjects design. Dependent variables included RT (s), error rate (%), Raw NASA-TLX (RTLX), and System Usability Scale (SUS). Outliers (8.5% of 810 trials) were removed due to abnormally slow RTs. After consent and calibration, participants completed a 40-trial practice block (SINGLE), followed by 30 experimental trials per condition with rest breaks and questionnaires between blocks.

4 Results

Table 1 summarizes descriptive statistics and repeated-measures ANOVA results. RT data were log-transformed to correct skew, and Greenhouse-Geisser corrections were applied where sphericity was violated. All other metrics met normality assumptions ($p > .05$).

One-way repeated-measures ANOVAs revealed no statistically significant main effect of visual cue type across any dependent variables ($p > .05$).

Table 1: Descriptive Statistics and ANOVA Results by Visual Cue Type

	SINGLE		DIRECTION		STOPLIGHT		ANOVA		
Measure	M	SD	M	SD	M	SD	F	p	η_p^2
Reaction Time (s)	1.53	0.58	1.63	0.55	1.77	0.91	0.57*	.577	.066
Error Rate (%)	15.36	7.81	17.27	11.54	16.07	6.04	0.15	.490	.070
NASA-TLX (RTLX)	35.28	24.29	36.67	24.14	43.79	22.42	2.04	.160	.200
SUS Score	83.61	22.33	76.39	28.09	77.50	21.36	0.47	.635	.055

Note. *Degrees of freedom for RT are (1.16, 9.3) due to Greenhouse-Geisser correction. All others are (2, 16).

4.1 Qualitative Feedback

Participants noted that DIRECTION and STOPLIGHT caused visual distraction. DIRECTION induced hesitation due to spatial arrow offsets, while STOPLIGHT overwhelmed users who misinterpreted the red target as a warning not to press. SINGLE was praised for targeting speed and simplicity, though participants noted STOPLIGHT could aid rhythm if visual sequences were refined. Crucially, feedback highlighted that precise physical calibration is essential: slight spatial offsets made target identification difficult.

5 Conclusion and Future Work

This study investigated whether added visual cue complexity is beneficial for AR-guided piano playing. While statistical power was limited (N = 9), descriptive trends indicate that multi-target previews (STOPLIGHT) increase cognitive load during rapid key-pressing. Conversely, minimalist single-target cues (SINGLE) enabled faster selection, though spatial offset in DIRECTION increased errors. Ultimately, these findings suggest that for immediate, sight-reading-style tasks in AR, minimalist visual guidance is currently more effective than complex predictive cues, though predictive designs may still hold value for rhythmic training if implemented carefully. Several key methodological limitations must be considered when interpreting these outcomes. First, an uncounterbalanced 40-trial practice block exclusively utilized the SINGLE cue, which effectively gave this baseline condition an unfair familiarity advantage before measurement. Second, the participant pool mixed total novices with highly experienced pianists (up to 16 years of experience). Averaging these extremes introduced high performance variance ($\eta_p^2 = .200$ for RTLX) and obscured how prior musical expertise changes how a user relies on AR cues. Finally, the evaluation utilized random note generation rather than actual musical pieces. As real music relies on predictable coherency (e.g., scales, chords, and rhythmic structures), a random-note task does not fully ascertain the utility of predictive cues in a genuine music education context. Future work will evaluate these techniques using actual musical pieces to better simulate the progression of learning a song. Furthermore, subsequent studies will use counterbalanced practice blocks, higher-precision marker-based tracking, and larger participant samples stratified by musical background to isolate the effects of prior expertise.

Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) Alliance Grant.

References

[1] Vivienne Amm, Krishnan Chandran, Lars Engeln, and Matthew McGinity. 2024. Mixed Reality Strategies for Piano Education. *Frontiers in Virtual Reality* 5 (10 2024). doi:10.3389/frvir.2024.1397154

[2] Minya Cai, Muhammad Alfian Amrizal, Toru Abe, and Takuo Suganuma. 2019. Design and Implementation of AR-Supported System for Piano Learning. 49–50. doi:10.1109/GCCE46687.2019.9015530

[3] Feng Huang, Yu Zhou, Yao Yu, Ziqiang Wang, and Sidan Du. 2011. Piano AR: A Markerless Augmented Reality Based Piano Teaching System, Vol. 2. 47 – 52. doi:10.1109/IHMSC.2011.82

[4] Leon Pietschmann, Michel Schimpf, Zhu-Tian Chen, Hanspeter Pfister, and Thomas Bohné. 2025. Enhancing User Performance and Human Factors through Visual Guidance in AR Assembly Tasks. 1–8. doi:10.1145/3706599.3720094

[5] Liam Rigby, Burkhard Wünsche, and Alex Shaw. 2020. piARno - An Augmented Reality Piano Tutor. 481–491. doi:10.1145/3441000.3441039

[6] Andra Simion, Adrian Iftene, and Daniela Gîfu. 2021. An Augmented Reality Piano Learning Tool. doi:10.37789/rochi.2021.1.1.21

[7] Benjamin Volmer, James Baumeister, Stewart Von Itzstein, Matthias Schlesewsky, Ina Bornkessel-Schlesewsky, and Bruce H. Thomas. 2023. Event Related Brain Responses Reveal the Impact of Spatial Augmented Reality Predictive Cues on Mental Effort. *IEEE Transactions on Visualization and Computer Graphics* 29, 12 (2023), 4990–5007. doi:10.1109/TVCG.2022.3197810

[8] Yuchong Zhang, Adam Nowak, Andrzej Romanowski, and Morten Fjeld. 2022. An Initial Exploration of Visual Cues in Head-Mounted Display Augmented Reality for Book Searching. In *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia* (Lisbon, Portugal) *(MUM '22)*. Association for Computing Machinery, New York, NY, USA, 273–275. doi:10.1145/3568444.3570600