https://doi.org/10.2352/EI.2026.38.13.ERVR-192

Actions can teach better than words

Michael Wilson*, Vinh Le *, Sergiu M. Dascalu;*, Frederick C. Harris, Jr.
Computer Science and Engineering, University of Nevada, Reno, Nevada, USA

Abstract

Monolingual learners frequently encounter barriers to language acquisition ranging from financial constraints to a lack of situational confidence. Virtual Reality (VR) offers a promising solution by providing a "safe" digital environment for immersive learning experiences. This paper evaluates a comparative study between two distinct delivery methods of a language lecture within VR: a traditional video presentation and a 3D-modeled experience utilizing consumer-grade motion capture hardware. In addition, this paper provides a solution with cheap consumer motion capture hardware, addressing the financial block above, to create educational content. Overall, results were mixed; while both the video and motion capture versions yielded positive engagement, the video format demonstrated a more significant quantitative increase in results. As part of the evaluation, we analyze the performance of low-cost motion capture in educational content creation and propose design iterations to better isolate the variables influencing these learning gains.

Introduction

Although traditional digital learning tools have evolved into sophisticated versions of the textbook—integrating spaced repetition systems (SRS), voice-recognition quizzes, and static video, they often fall short of providing a truly interactive environment. These platforms provide a structured path for vocabulary and grammar, yet they rarely offer the immersive challenges necessary to force active communication and problem-solving. For many monolingual speakers, the struggle to learn a second language is rooted in a lack of situational confidence, as much as it is in financial constraints. Without a low-stakes, safe environment to practice, students often remain hesitant to engage, highlighting the need for educational tools that move beyond the screen and into a more experiential space.

To address these gaps, this paper introduces a Virtual Reality application built specifically to facilitate language learning. We demonstrate that high-quality educational content can be produced using accessible, consumer-grade VR equipment for both actor motion capture and student playback, effectively lowering the financial entry barrier for immersive tech. We provide an evaluation of this approach through a study comparing a video and motion capture version of a lecture.

The scientific inquiry of this paper is guided by two central research questions:

Question 1: Does a 3D motion-captured puppet animation in VR offer pedagogical advantage over a standard flat video presentation for teaching a language?

Question 2: Does consumer VR hardware have enough capability to author motion-captured content useful for improving confidence in peoples' ability to learn a language?

These questions are explored through a user study involving the language Toki Pona; chosen for its minimal 120-word vocabulary, it serves as an ideal baseline for the experiment. This context allows participants to experience the full arc of learning a new language in a condensed time frame, providing a clear window into how the delivery medium influences their perception of success and their willingness to continue their studies.

The remainder of this paper is structured as follows. Section discusses an overview of SRS and similar approaches to the system proposed in this paper. Section covers the design of the software approach. Section covers the methodology of the studies performed as part of this work. Section provides the results and Section details a discussion of the insights gained from both studies. Finally, Section wraps up the paper with the conclusions and work to come.

Background and Related Work

The landscape of self-paced language acquisition offers the idea of "comprehensible input," a theory suggesting that learners acquire language most effectively when exposed to target-material that is contextualized and understandable [10]. This input-driven philosophy is epitomized by foundational works like Lingua Latina per se Illustrata: Familia Romana, which utilizes a self-contained methodology requiring no bridge language to teach Latin [8]. Recent practical applications of this theory have surfaced within the Toki Pona community, notably through the work of practitioners like Jan Telakoman, who utilizes conversational podcasts to demonstrate the rapid acquisition of the language through consistent, high-quality input [11].

Complementing input-heavy methods are Spaced Repetition Systems (SRS), which serve as an evolution of the traditional textbook by optimizing memory retention. Systems such as WaniKani and Bunpro use personalized server-based tracking to prompt student recall at the exact moment of potential decay, while older audio-based systems such as Pimsleur apply a more static version of this logic [2]. These tools meet the logistical needs of modern students, yet often lack the immersive elements required to simulate real-world communication challenges.

In this context, Toki Pona serves as an ideal linguistic vehicle to testing immersive technologies. Since the publication of the official language specifications in 2014 [5], the language has moved from technical linguistic definitions to structured speaking guides [4]. Because of its minimal 120-word vocabulary, Toki Pona allows a student to bypass the extensive time commitments of natural languages. This small scope for learning a language also provides an opportunity for research. It is suitable for use in a controlled environment to evaluate whether Virtual Reality and consumer-grade motion capture can bolster a learner's situational confidence.

Figure 1. *Experiment setup with Quest Pro headset*

Related Works

Computer-Aided Language Learning (CALL) is a type of computer-aided instruction. Since the 1960s, computers have been used for assisting language learning. The early programs were focused on language drilling and practice, where students would be given forms to use and repeat. In the following decades, the idea of teaching forms implicitly rather than explicitly began to gain traction. Students should be able to produce new phrases instead of repeating existing exact patterns. A focus would be on what students would do with one another while working on the machine, rather than what work was done with the machine. [12]

VR/AR being used to aid language learning is not a new concept, and while narrow of a field, it has had its fair share of research over the years [9, 3, 6]. However, emphasizing the belief mentioned above, one particular work stands out in the social VR space by the name of Hololingo! This apparatus operates as an application within the VRChat platform and focuses on teaching German. It uses a multiplayer experience with several stages bound to different learning or social goals to achieve that belief in interaction. [1].

Experiment Considerations & Software Design

To evaluate the impact of immersion on language acquisition, this research identifies three distinct modes of Virtual Reality engagement: non-interactive motion capture playback, semi-interactive ”paused” experiences where learners take specific actions, and fully autonomous environments where characters react to the user in real-time. This paper focuses on the first level of immersion, utilizing a classroom-based narrative. This scenario places a participant in a university setting where a peer provides a targeted introduction to the Toki Pona language. The experience is designed as an eight-minute session, concluding with the participant transitioning out of the virtual environment.

The primary objective of this experiment is to determine if participants can acquire the fundamental components of Toki Pona, specifically a subset of 20 words, to describe their immediate surroundings. While the lesson introduces core grammatical structures, we recognized that the eight-minute duration provides a density of content that may exceed what a student can fully absorb in a single sitting. To maintain experimental control and minimize technical friction for the initial study, the session requires no active interaction from the participant. Efficacy is measured through a comparative analysis of participant performance between pre-test and a post-test questionnaires. An example of our physical experiment setup is presented in Fig. 1.

Animation within the environment is driven primarily by motion capture, which was conducted by the authors. To address the goal of using cost-effective resources, animation data is gathered using the Meta Quest 3. This data is recorded directly within the virtual environment using Shadermotion, a tool that records performance data from within VRChat. This eases the process of recording for the actor, as their motion data is mapped to same context as in the final production.

The experience is developed atop the Unity-based VRChat platform. As a leading social VR ecosystem, VRChat provides a robust, cross-platform foundation compatible with both Windows and Android-based VR headsets. Utilizing this platform offers significant development advantages, including a sophisticated inverse kinematics (IK) system for natural avatar movement, established multiplayer networking, and accessibility features. By building within this established framework, the project leverages a stable environment for evaluating how immersive ”puppet” animations compare to traditional video formats.’

Methodology

To investigate the impact of immersive motion capture on language acquisition, this study employed a between-subjects design comparing two distinct delivery modes within a Virtual Reality environment. This section details the demographics of the participants, the experimental environment, and the systematic procedure used to isolate the effects of the presentation style on the confidence and retention of the learner.

Participants

A total of 36 participants were recruited from the University of Nevada, Reno, primarily through targeted email campaigns and word-of-mouth. Eligibility was restricted to individuals over the age of 18 who were willing and able to wear a VR headset for the duration of the study. The cohort consisted of university students and faculty members, representing a range of prior experiences with immersive technology. Participation was entirely voluntary, and no financial compensation was provided.

Experimental Setup

The study was conducted in a controlled laboratory setting on the fourth floor of the William N. Pennington Engineering Building (WPEB) at the University of Nevada, Reno. To ensure hardware consistency across all trials, every participant utilized the Meta Quest Pro VR headset. This device was selected for its high-fidelity display and comfort, minimizing the potential for physical distraction or hardware-induced discomfort during the educational session. Despite its comfort, wearing a VR headset

Figure 2. *Immersive version of the experiment*

Figure 3. *Traditional 2D version of the experiment*

can feel unnatural. Both groups in the study wear the headset to eliminate potential variables that are unique to wearing the 800 gram device on one's head. This can help exclude differentiating potential factors such as VR sickness, neck pain, claustrophobic feelings, etc.

Experimental Procedure

Upon arrival at the laboratory, participants were briefed on the study's objectives and provided with a formal consent form. Following consent, the session followed a structured four-phase workflow.

Participants first completed a pre-study survey to establish a baseline. This included demographic data, frequency of VR usage, and qualitative opinions on VR comfort using Likert scale questions. To ensure a clear delta in learning, participants also took a preliminary quiz on Toki Pona. As the study utilized a niche language with only 120 words, it was expected that participants unfamiliar with the language would score zero, allowing an accurate measurement of knowledge gain.

After the assessment, the facilitators provided a brief orientation on the Quest Pro hardware. The participants were then assisted in fitting the headset to ensure optimal visual clarity and physical comfort before the virtual lecture began. Participants were randomly assigned to one of two groups that mapped to the traditional video or the motion-captured version. Both groups experienced an 8-minute Toki Pona lesson set within a virtual classroom, but the delivery method varied by group.

For the first group, participants viewed the lecture as a flat 2D video projected onto a wall within the virtual environment, as seen in 3. This served as a digital representation of traditional, non-immersive educational media. As part of the second group, participants observed the same lecture content, but instead of a screen, they interacted with a life-sized 3D avatar animated via the previously recorded consumer-grade motion capture. This provided a "puppet" presence intended to test the efficacy of spatialized human-like movement, as seen in 2

Immediately following the conclusion of the 8-minute session, participants removed the headset to complete a post-quiz and a final qualitative survey. This allowed for a direct comparison of retention results and a final assessment of the level of confidence of the learner after the intervention.

The primary metric of success was the improvement in score between the pre-quiz and post-quiz, specifically focused on the 20 target words and basic grammar structures introduced during the session. Identical audio and instructional content between both the video and avatar groups is maintained. The primary focus of this study was evaluating the delivery method: a traditional flat video versus a 3D animation. The goal was to evaluate if a 3D immersive animation has stronger impact on student information retention than a flat video.

Results

The study yielded a complex and encouraging set of results in 36 participants. Although the primary objective was to evaluate the efficacy of VR delivery modes, the data also provided interesting insights and the logistical nuances of immersive educational content.

Quantitative Quiz Performance

As a whole, the design of the content showed to be largely positive, with nearly every participant demonstrating measurable improvement between the pre-test and post-test, in both groups. We categorized the baseline knowledge into nine "buckets" to track progress relative to starting proficiency, as shown in Table 1. Notably, of the 18 participants who began with zero knowledge of Toki Pona, nearly all showed improvement, achieving a mean score of 5.77 and a median of 6 (out of 8). Only two participants in this group scored below 50% in the post-test. As seen in 4, participants scored much better in the post-quiz for each question.

For those with marginal prior exposure—scoring a 1 on the pre-test (n=8), the results mirrored this success, with a median post-test score of 6 and all users surpassing the 50% mark. While those starting with higher baseline scores (3 to 4) continued to show growth, participants entering with near-perfect scores (6 to 8) maintained their level, indicating that the 8-minute module successfully conveyed many of the target 20 words and basic grammar to the vast majority of the cohort regardless of their starting point.

Across the board with every question there were better outcomes in the Flat Video groups as compared to the Puppet Animation groups, as seen in 5.

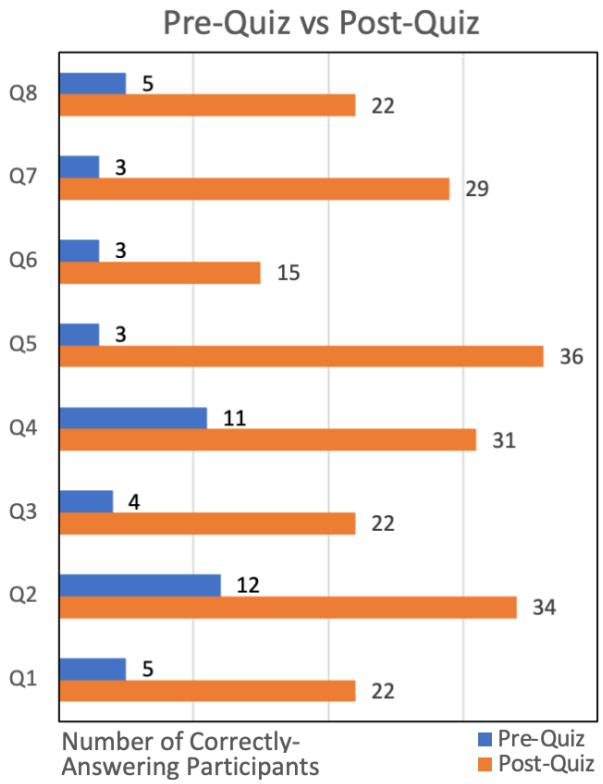


Figure 4. *Pre and Post Quiz Results*

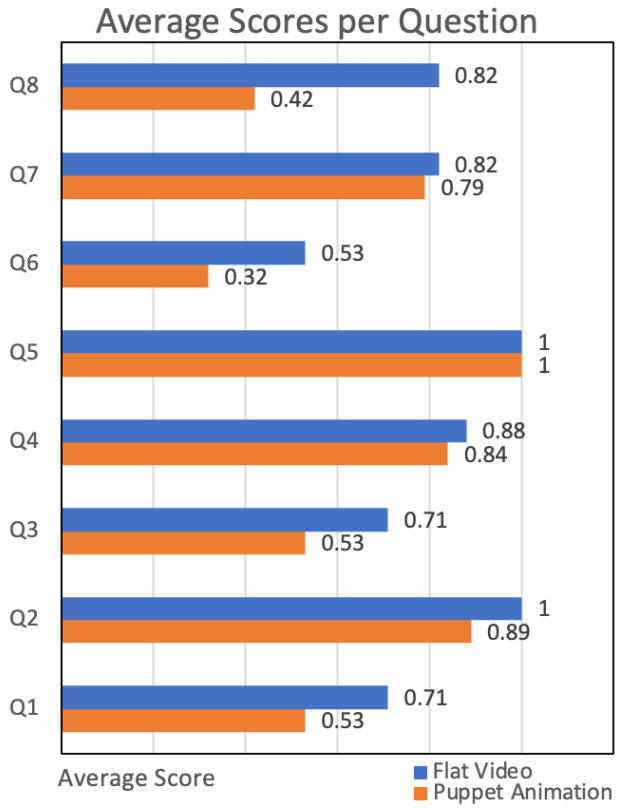


Figure 5. *Post-Quiz results by group*

Attitudinal and Comfort Metrics

The Likert scale data revealed an upward trend in learner confidence and interest, presented in Table 2. The most significant gains were observed in Question 1 (confidence in language study), which rose from a mean of 3.57 to 4.09, and Question 4 (perceived utility of Toki Pona), which saw a substantial increase of 0.94. Physical comfort remained high throughout the study; despite the use of the Meta Quest Pro, the average comfort rating improved from 3.74 to 4.14, suggesting that the immersive experience did not negatively impact the physical well-being of the participants.

However, a methodological complication was identified regarding Questions 4 and 6, which inquired about Toki Pona before the lesson had occurred. Participants unfamiliar with the language required clarification during the pre-survey, and were instructed to select a neutral ”3.” This highlights a specific area for refinement in the future: ensuring that definitions and context are provided consistently before measuring attitudinal shifts.

Qualitative Feedback and Pedagogical Implications

Participant comments provided critical insights into their perception of the virtual environment. Several users specifically noted that the use of anime-style avatars reduced the classroom-related anxiety they typically experience in traditional settings. While 10 of the 36 responses cited minor hardware issues such as weight or visual blurriness, the majority praised the Quest Pro’s ergonomics, particularly its compatibility with prescription glasses.

Key suggestions for future iterations centered on increas-

Bucket	n	Average Post-Test Score	Median Post-Test Score	Improvement Notes
0	18	5.77	6	100% showed improvement; only 2 scored below 50%.
1	8	5.75	6	100% showed improvement; all scored $\geq$ 50%.
2	2	6	6	Both participants tripled their initial score.
3	4	6	6.5	3 of 4 participants significantly improved.
4	2	5.5	5.5	Both participants showed moderate gains.
6	1	6	6	Participant maintained a high baseline score.
8	1	8	8	Participant maintained a perfect score.

Table 1: Buckets Scores

Survey Question (Likert 1-5)	Pre-Survey Mean	Post-Survey Mean	Improvement
Q1: Confidence in language study	3.57	4.09	+0.52
Q2: Desire to learn another language	4	4.43	+0.43
Q3: Enjoyment of language learning	4.14	4.57	+0.43
Q4: Toki Pona as a useful starter	3.31	4.26	+0.95
Q5: Physical comfort in VR	3.74	4.14	+0.40
Q6: Desire to learn Toki Pona	3.74	4.03	+0.29
Q7: Desire to learn another language	4.11	4.37	+0.26
Q8: Desire to go to another country	4.43	4.43	+0.00

Table 2: Survey Results

ing interactivity. Participants expressed a desire for a "running glossary" within the virtual space and a recap phase at the end of the lecture. One significant takeaway for a future iteration is the request for a pre-check period to adjust the virtual environment (e.g., desk height and brightness) before the playback begins. These qualitative findings suggest that while the motion-captured "puppet" and video modes are both effective, the true potential of the application lies in its ability to mitigate social barriers to learning.

Discussion

The preliminary results of this study offer a nuanced perspective on the efficacy of immersive motion capture compared to traditional video formats. While both groups demonstrated significant learning gains, the results for Research Question 1—whether a 3D puppet animation outperforms flat video—were mixed. Counterintuitively, the flat video group exhibited slightly higher performance metrics. However, for Research Question 2, the data strongly suggests that consumer-grade VR hardware can indeed foster language learning and bolster student confidence. The fact that a successful educational experience was authored and delivered on a budget of under $500 USD (the price of a Meta Quest 3, used for the motion capture process) indicates that VR is a viable, low-cost solution for democratizing immersive education.

Analysis of Experimental Bias

The unexpected performance of the flat video group may be attributed to two primary confounding variables: sampling bias and the "novelty effect." The study was conducted over five days, with the flat video group largely concentrated in the first forty-eight hours. These early participants displayed a higher degree of baseline enthusiasm and prior familiarity with the concept of the language. Furthermore, for participants with limited VR experience, the presence of a life-sized, animated 3D avatar may have acted as a cognitive distraction rather than a pedagogical aid. In future iterations, we will incorporate specific demographic queries regarding the participants' native and secondary languages. Identifying non-monolingual speakers is essential, as polyglots may possess a higher cognitive threshold for acquiring a simplified language like Toki Pona, which could mask the efficacy of the VR delivery for our primary target audience.

Experimental Apparatus & Social VR

To mitigate the bias caused by hardware unfamiliarity, we propose transitioning future studies into established Social VR ecosystems such as VRChat or Resonite. Recruiting "native" VR users allows the study to focus on the instructional content rather than the user's adjustment to the interface. This transition also enables the introduction of interactive elements that were previously omitted for simplicity. In a Social VR context, the eight-minute lecture could be followed by a guided exploration phase. For example, interactive posters within the virtual lecture hall could serve as communicative tasks; when a student approaches a diagram of grammatical particles, the lecturer could offer a contextual explanation, reinforcing the lecture through spatial discovery.

Feedback Loops and Recasting

A critical component of Task-Based Language Teaching (TBLT) is the presence of "negative feedback," specifically through a process known as recasting [7]. Recasting occurs when an instructor subtly corrects a learner's error by repeating the sentence back in its correct form within the flow of conversation. While real-time, autonomous recasting is technically challenging for a pre-recorded VR experience, we propose a self-administered feedback loop. Given that Toki Pona features a highly consistent orthography and simple stress patterns, students can be prompted with call-and-response exercises. By playing back the student's own voice immediately followed by the correct pronunciation, the system can provide a "gentle" corrective feedback.

Reducing Variables

The findings from this initial study show the necessity of reducing additional variables. Possible distraction from users' non-familiarity with the hardware is a possible interpretation of the results.

Conclusions

Overall, this paper details a study conducted in VR that compares the efficacy of a 3D motion-captured animation over a 2D video when used as a medium for language learning. By utilizing the minimalist vocabulary of Toki Pona, the study achieved a promising improvement rate among absolute beginners, with significant gains in both linguistic retention and learner confidence. Although the performance of the flat video group compared to the 3D avatar suggests that the "novelty effect" of VR can initially act as a cognitive distractor for those unfamiliar with the hard-

ware, the qualitative feedback underscores VR's unique ability to mitigate classroom anxiety through "safe" digital environments. Ultimately, this work contributes both a cost-effective strategy for immersive content creation and a robust framework for isolating the specific variables that lead to successful digital language acquisition.

Future Work

Subsequent research will expand the current experiment into a comprehensive six-module curriculum designed to guide learners through the entire 120-word Toki Pona lexicon. This narrative-driven series simulates a journey to a foreign country, progressing from passive classroom instruction to high-stakes professional interaction. Each module is capped at a fixed 24-minute duration to ensure experimental consistency, yet features "adaptive density" where more proficient learners receive advanced prompts within the same time frame. These future modules are designed to be replayable, allowing for longitudinal studies on how iterative exposure to immersive "fail-safe" environments affects long-term retention and situational confidence.

Proposed Learning Modules:

- Module 1 (Classroom): Fundamental grammar and vocabulary using non-interactive motion capture.
- Module 2 (Backpacking): Casual Self-paced conversation and environmental description.
- Module 3 (Campsite): Functional command-based tasks involving "fail-safe" NPC assistance.
- Module 4 (Travel): Immersive small talk and social interaction during a simulated flight.
- Module 5 (Transit): Navigation and comprehension of the written form of a language in a train station.
- Module 6 (Conference): Synthesis of all concepts through a professional presentation and Q&A

Acknowledgments

This material is based in part upon work supported by the National Science Foundation under grants IIS-2202640 and OIA-2148788. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Timo Ahlers et al. "Hololingo! - A Game-Based Social Virtual Reality Application for Foreign Language Tandem Learning". In: *DELFI 2021*. Bonn: Gesellschaft für Informatik e.V., 2021, pp. 37–48.

[2] Marc Butler. "Immersive Japanese Language Learning WEB Application Using Spaced Repitition, Active Recall, and an Artificial Intelligent Conversational Chat Agent both in Voice and in Text". `https://knowledge.e.southern.edu/mscs_reports/11/`. MA thesis. Southern Adventist University, School of Computing, Apr. 2024.

[3] Xinyi Huang et al. "A systematic review of AR and VR enhanced language learning". In: *Sustainability* 13.9 (2021), p. 4639.

[4] Sonja Lang. *lipu sona pona (toki pona course)*. `https://lipu-sona.pona.la/index.html`. Last Accessed May 15, 2024.

[5] Sonja Lang. *Toki Pona (official site)*. `https://tokipona.org/`. Last Accessed May 15, 2024.

[6] Tsun-Ju Lin and Yu-Ju Lan. "Language learning in virtual reality environments: Past, present, and future". In: *Journal of Educational Technology & Society* 18.4 (2015), pp. 486–497.

[7] Mike Long. *Second Language Acquisition and Task-Based Language Teaching*. `https://www.wiley.com/en-us/Second+Language+Acquisition+and+Task-Based+Language+Teaching-p-9780470658949`. Wiley-Blackwell, 2014.

[8] Hans H. Ørberg. *Lingua Latina: Pars I: Familia Romana*. Second. `https://hackettpublishing.com/lingua-latina-pars-i-familia-romana-full-color-edition`. Hackett Publishing, 2011.

[9] Antigoni Parmaxi. "Virtual reality in language learning: A systematic review and implications for research and practice". In: *Interactive learning environments* 31.1 (2023), pp. 172–184.

[10] Robert Patrick. "Comprehensible Input and Krashen's Theory". In: *Journal of Classics Teaching* 20 (39 2019), pp. 37–44.

[11] Jan Telakoman. *30-Day Comprehensible Input Challenge*. `https://30dcic.my.canva.site/`. Last Accessed May 15, 2024.

[12] Mark Warschauer and Deborah Healey. "Computers and language learning: an overview". In: *Language Teaching* (1998).

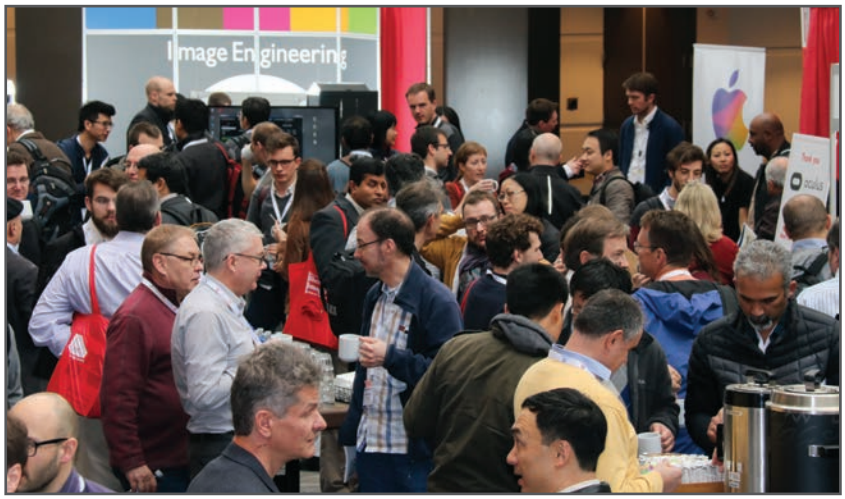
Author Biography

Michael Wilson specializes in Virtual Reality and Human-Computer Interaction (HCI). They earned their Bachelor's degree in 2017 and Master's Degree in 2025 from the University of Nevada, Reno, where they worked as a VR Development Specialist for the University Libraries. Michael is passionate about languages' ability to connect people, especially accessible constructed languages like Toki Pona. Michael is currently pursuing a PhD at the University of Otago in Dunedin, NZ.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 1947 IS&T 2022 imaging.org