MundoVR: A Systems Engineering Approach to Gamified Second Language Acquisition with AI and VR

Ashley NAKA (819029)
Jennifer AWOUNOU (819021)
Azamkhon Khudoyberdiev (819025)

Abstract

Motivation Traditional language learning apps rely on rote memorization and fail to prepare learners for spontaneous real-world dialogue. Recent advances in LLMs and voice processing enable personalized, interactive systems that simulate authentic human conversation.

Problem The challenge is designing a robust, real-time system delivering high-quality, low-latency conversational experiences in immersive VR while remaining cost-effective and scalable. Key trade-offs involve balancing on-device and cloud computational resources, maintaining low latency for VR immersion, and ensuring system resilience under variable network conditions.

Solution We propose MundoVR, a hybrid AI architecture combining lightweight on-device AI (NLU, STT, embeddings) for rapid interactions with cloud-based LLMs for complex dialogues. The platform integrates STT, TTS, and phoneme-level analysis within a scalable microservices backend (Golang, Kubernetes), optimized for VR real-time constraints.

Contribution This work specifies a hybrid software architecture for real-time intelligent educational applications, addressing critical trade-offs between latency, computational cost, and AI-driven interaction quality through formal specification using UML diagrams and architectural patterns.

System Specification Approach

We specify the MundoVR system using a multi-view architectural methodology, ensuring a comprehensive and standardized description.

- **Modeling Language:** UML 2.5 notation following ISO/IEC/IEEE 42010 architectural description standards.
- **Views:** We use a (1) Context Diagram to define system boundaries and external actors, a (2) Deployment Diagram to map components to infrastructure, and a (3) Sequence Diagram to detail key interaction flows.
- **Decision Documentation:** Architecture Decision Records (ADRs) are used to capture key trade-offs, such as latency vs. quality, cost vs. performance, and on-device vs. cloud processing.

This approach ensures the systematic documentation of architectural choices critical to the hybrid AI system.

Requirements Engineering

Domain and Stakeholders

MundoVR operates at the intersection of Second Language Acquisition (SLA), Artificial Intelligence (AI), and Virtual Reality (VR), providing a gamified, immersive environment for conversational practice with AI-driven virtual characters.

User Personas

- **High School Student (Leila):** Preparing for language exams; needs structured lessons and grammar drills integrated into conversation.
- **Business Professional (Mark):** Learning industry-specific jargon for international meetings; focuses on formal communication and negotiation scenarios.
- **Casual Tourist (Alex):** Learning basic conversational phrases for travel; needs practical scenarios like ordering food and asking directions.

System Context and Interfaces

MundoVR employs a hybrid architecture that distributes workloads between on-device processing and cloud services to balance performance and cost. The system defines four key interfaces:

- **User-VR Headset:** Voice commands and physical interactions within the VR environment.
- **VR Headset-AI SDK:** Low-latency local connection for simple interactions (e.g., command acknowledgment).
- **VR Headset-Cloud Platform:** Secure HTTPS API calls for complex conversational processing.
- **Platform-Cloud AI:** Backend communication with LLM, STT, and TTS services.

Use Cases and Scenarios

The following scenarios are derived from the user stories to illustrate key system functionalities.

Use Cases

Five key scenarios illustrate system functionality, derived from user personas:

- **US1 (Exam Preparation):** High school student practices debate about environmental policies with an AI tutor to master vocabulary and arguments for oral exams.
- **US2 (Business Negotiation):** Business professional simulates contract negotiations to practice formal language and industry-specific terminology.
- **US3 (Travel Practice):** Casual tourist role-plays ordering meals at virtual restaurants to build confidence for travel.
- **US4 (Pronunciation Clinic):** System identifies specific pronunciation errors and provides targeted drills for accent improvement.
- **US5 (Adaptive Role-play):** AI characters react dynamically to learner choices, creating authentic and engaging conversations.

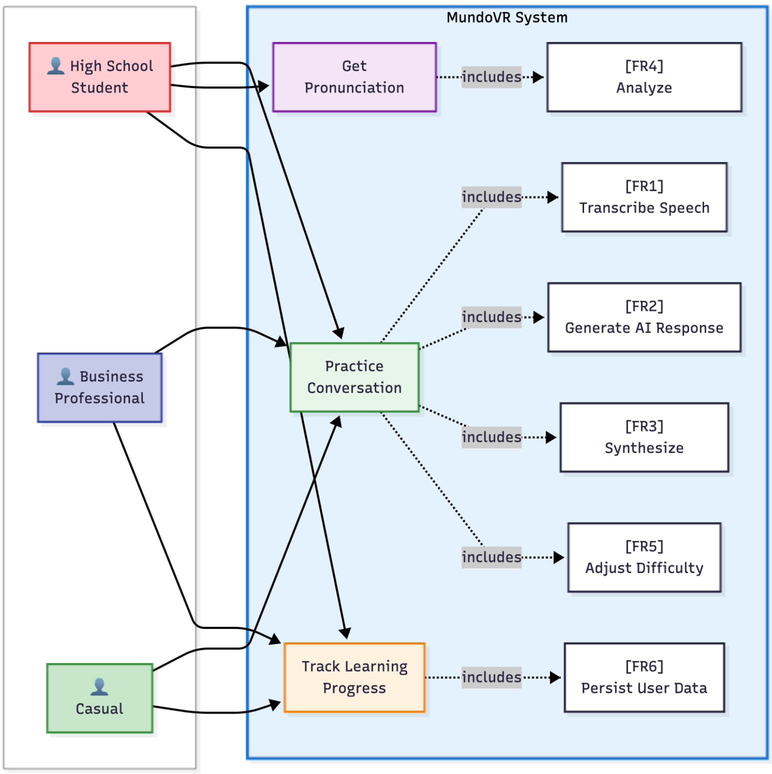

Figure 1: Use case diagram showing actor interactions with system functional requirements.

Functional Requirements

- **FR1:** The system shall transcribe user's spoken input into text in real-time.
- **FR2:** The system shall generate contextually relevant responses from an AI character.
- **FR3:** The system shall synthesize text responses into natural-sounding speech.
- **FR4:** The system shall analyze user's speech for pronunciation accuracy.

- **FR5:** The system shall adjust dialogue complexity based on user performance.
- **FR6:** The system shall persist user profiles, progress, and conversation history.

Non-Functional Requirements

- **NFR1 (Latency):** Two-tier response time for VR immersion:
  - *Tier 1:* On-device AI SDK responses ¡ 200ms
  - *Tier 2:* Full cloud loop (STT $\rightarrow$ LLM $\rightarrow$ TTS) ¡ 1.5 seconds
- **NFR2 (Scalability):** Support 10,000 concurrent users. Graceful degradation under peak load increases response time by up to 50% while prioritizing core conversation features.
- **NFR3 (Cost-Effectiveness):** At least 40% of interactions handled by on-device AI SDK to minimize cloud processing costs during typical 15-minute sessions.

System Architecture

The hybrid architecture distributes workloads between on-device and cloud components to meet latency, scalability, and cost requirements.

Context Diagram

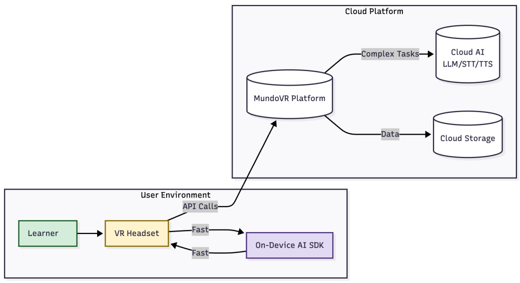

Figure 2: System context showing hybrid on-device/cloud architecture.

Service Architecture

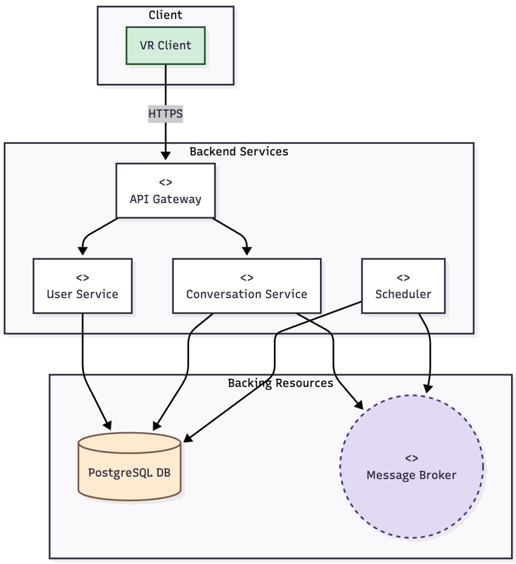

Figure 3: Microservice components with data stores and message brokers.

Deployment Architecture

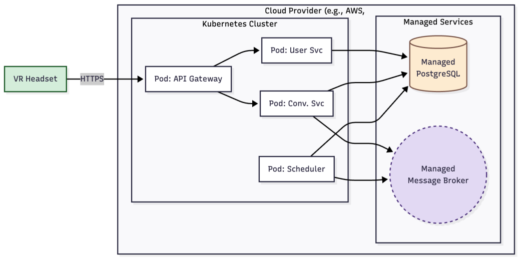

Figure 4: Containerized services in Kubernetes cluster with managed cloud resources.

Literature Review

VR and Immersive Technologies for Language Learning

- **Schorr, I., et al. (2024)** [1] — Systematic review of 40 studies (2016-2023) validating VR for contextual language learning and vocabulary acquisition.
- **Oxford University Press (2025)** [2] — Pedagogical frameworks for VR-based SLA, informing scenario design and dialogue adaptation.
- **Frontiers in Virtual Reality (2025)** [3] — AR vs. VR comparison demonstrating enhanced retention in VR, supporting our VR-focused approach.

AI Integration in Educational VR

- **Adithya, T. G., et al. (2024)** [4] — GPT-based AI tutoring with Unity 3D VR, providing technical precedent for our hybrid architecture.
- **Godwin-Jones, B. (2024)** [5] — Analysis of AI/VR convergence highlighting on-device vs. cloud processing balance.
- **Zhu, Z., et al. (2025)** [6] — CHI 2025 study on LLM-powered pedagogical agents in VR, validating adaptive role-switching and dialogue management.
- **Vallance, M. (2024)** [7] — AI-enabled NPCs for educational VR, informing on-device AI SDK and cloud LLM integration strategy.
- **IEEE VRW (2025)** [8] — Parallel implementation of AI-driven Japanese learning in VR with contextualized conversations and adaptive task complexity.

Conversational AI and Chatbots in Education

- **Kuhail, M. A., et al. (2023)** [9] — Systematic review of 36 educational chatbot implementations showing improved learning outcomes with personalized approaches, informing our AI character design.
- **Velazquez-Garcia, L., et al. (2024)** [10] — AI integration in educational gamification, providing insights for adaptive content delivery and engagement mechanics.

Second Language Acquisition Theory

- **Krashen, S. D. (1982)** [11] — Comprehensible Input hypothesis (i+1) informs dialogue difficulty adaptation; Affective Filter hypothesis guides low-pressure environment design.
- **Long, M. H. (1996)** [12] — Interaction Hypothesis validates conversational interaction focus and negotiation of meaning in dialogue design.
- **Swain, M. (1995)** [13] — Output Hypothesis supports active speech production and pronunciation analysis (FR4).
- **Ellis, R. (2008)** [14] — Ten principles of instructed SLA provide framework for scenario design and feedback mechanisms.
- **Gass, S. M. & Selinker, L. (2013)** [15] — Comprehensive SLA overview establishing theoretical foundation for input, interaction, and output integration.
- **VanPatten, B. (2017)** [16] — Processing Instruction theory guides structured input presentation for form-meaning connections.

References

[1] Isabel Schorr et al. "Foreign language learning using augmented reality environments: a systematic review". In: *Frontiers in Virtual Reality* 5 (2024), p. 1288824. DOI: 10.3389/frvir.2024.1288824. URL: https://doi.org/10.3389/frvir.2024.1288824.

[2] Oxford University Press. "Virtual Reality for Language Learning". In: *ELT Journal* (2025). DOI: 10.1093/elt/ccaf025. URL: https://academic.oup.com/eltj/advance-article/doi/10.1093/elt/ccaf025/8186115.

[3] Frontiers in Virtual Reality. "Analyzing and Comparing Augmented and Virtual Reality in Vocabulary Learning". In: *Frontiers in Virtual Reality* (2025). DOI: 10.3389/frvir.2025.1522380. URL: https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2025.1522380/full.

[4] Adithya T. G., Abhinavaram N., and Gowri Srinivasa. "Leveraging Virtual Reality and AI Tutoring for Language Learning: A Case Study of a Virtual Campus Environment with OpenAI GPT Integration with Unity 3D". In: *arXiv preprint arXiv:2411.12619* (2024). DOI: 10.48550/arXiv.2411.12619. URL: https://arxiv.org/abs/2411.12619.

[5] Bob Godwin-Jones. "AI and VR Converge: The Future of Language Learning in an Emerging XR Ecosystem". In: *Self-Published Research Report* (2024). URL: https://godwinjones.com/godwin-jones_vr_ai_merge.pdf.

[6] Zihao Zhu et al. "Exploring LLM-Powered Role and Action-Switching Pedagogical Agents for History Education in Virtual Reality". In: *CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Apr. 2025. DOI: 10.1145/3706598.3713109. URL: https://dl.acm.org/doi/10.1145/3706598.3713109.

[7] Michael Vallance. "Virtual Heritage & AI: Learning about 19th Century Japan with Isabella Bird". In: *ICEEL '23: Proceedings of the 2023 7th International Conference on Education and E-Learning*. May 2024, pp. 6–13. DOI: 10.1145/3637989.3638007. URL: https://dl.acm.org/doi/10.1145/3637989.3638007.

[8] "Immersive AI-Powered Language Learning Experience in Virtual Reality: A Gamified Environment for Japanese Learning". In: *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Saint Malo, France: IEEE, Mar. 2025. ISBN: 979-8-3315-1484-6. DOI: 10.1109/VRW66409.2025.00455. URL: https://ieeexplore.ieee.org/document/10972685.

[9] Mohammad Amin Kuhail et al. "Interacting with educational chatbots: A systematic review". In: *Education and Information Technologies* 28 (2023), pp. 973–1018. DOI: 10.1007/s10639-022-11177-3. URL: https://link.springer.com/article/10.1007/s10639-022-11177-3.

[10] Lydia Velazquez-Garcia et al. "Enhancing Educational Gamification through AI in Higher Education". In: *ICETC '24: Proceedings of the 2024 16th International Conference on Education Technology and Computers*. Sept. 2024, pp. 213–218. DOI: 10.1145/3702163.3702416. URL: https://dl.acm.org/doi/10.1145/3702163.3702416.

[11] Stephen D. Krashen. *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press, 1982. ISBN: 0-08-028628-3. URL: http://www.sdkrashen.com/content/books/principles_and_practice.pdf.

[12] Michael H. Long. *The Role of the Linguistic Environment in Second Language Acquisition*. Ed. by William C. Ritchie and Tej K. Bhatia. San Diego, CA: Academic Press, 1996, pp. 413–468.

[13] Merrill Swain. "Three Functions of Output in Second Language Learning". In: *Principles and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson* (1995). Ed. by Guy Cook and Barbara Seidlhofer, pp. 125–144.

[14] Rod Ellis. "Principles of Instructed Second Language Acquisition". In: *CAL Digest* (2008). URL: https://www.cal.org/resource/principles-of-instructed-second-language-acquisition/.

[15] Susan M. Gass and Larry Selinker. *Second Language Acquisition: An Introductory Course*. 4th. New York: Routledge, 2013. ISBN: 978-0-415-89488-4.

[16] Bill VanPatten. "Processing Instruction". In: *The Routledge Handbook of Instructed Second Language Acquisition* (2017). Ed. by Shawn Loewen and Masatoshi Sato, pp. 166–184.