XR SALENTO 2025
International Conference on eXtended Reality
June 17-20, 2025 – Otranto, Italy

AI-based Services to Support Language-Learning for Deaf and Hearing Individuals in Immersive XR Settings

N.D. Tantaroudas, A. J. McCracken, I. Karachalios, and E. Papatheou
Institute of Communication and Computer Systems, 9 Ir. Polytechneiou, Zografou, 15773, Greece, https://www.iccs.gr/el/
DASKALOS-APPS, Rue de l'abbe Griffon, 01960, Peronnas, France, https://daskalos-apps.com
Department of Water Resources and Environmental Engineering, School of Civil, Engineering, National Technical University of Athens, Athens, Greece, http://www.hydro.civil.ntua.gr/en/
Exeter Small-Scale Robotics Laboratory, Engineering Department, University of Exeter, Exeter, EX4 4QF, UK, https://engineering.exeter.ac.uk/
nikolaos.tantaroudas@iccs.gr

Introduction and Motivation

Traditional language-learning methods such as grammar-translation [1] and audio-lingual drills [2], often struggle with engagement, context relevance, and inclusivity, especially for deaf learners. Extended Reality (XR), combined with Artificial Intelligence (AI), offers new opportunities for immersive, interactive, and inclusive educational environments [3]. Tools like virtual avatars, speech-to-text, and text-to-sign translation can enhance inclusivity for both hearing and deaf users [4]. However, current systems often fail to support sign language, multilingual access, or reusability at scale [5]. INFINITY addresses these gaps by integrating modular AI services into XR platforms to support self-directed, equitable, and multilingual learning experiences.

Methodology and Design

The INFINITY platform features a modular, scalable architecture offering robust interoperability across XR applications via API-driven AI microservices. XR scenarios are pedagogically aligned, co-designed with educators and language specialists, focusing on communicative and task-based learning methods. Key services include:

- **Speech-to-text transcription** enabling real-time captions (Fig. 2).
- **Text-to-speech translation** facilitating multilingual spoken interactions (Fig. 3).
- **Text-to-sign translation** bridging spoken/written language and International Sign Language (ISL).
- **LLM-driven summarization** for immersive session review and personalized learner feedback (Fig. 4).

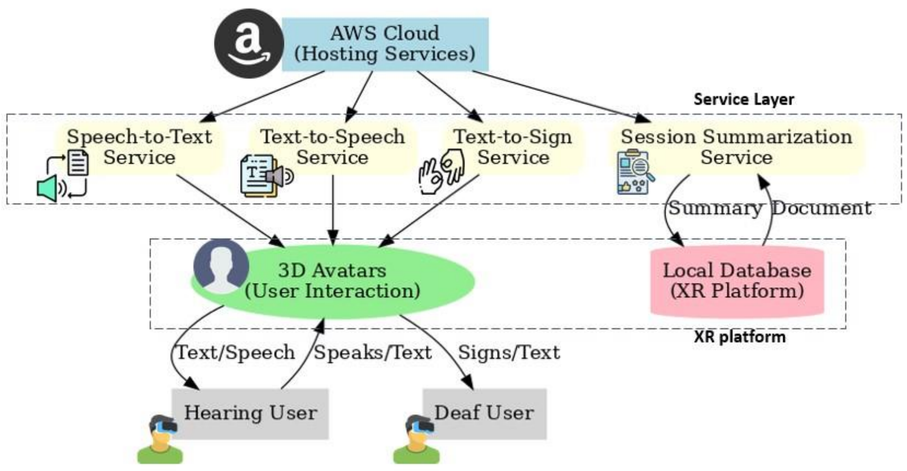

Figure 1: *INFINITY conceptual framework demonstrating integration of AI services (speech-to-text, text-to-speech, text-to-sign) with 3D avatars in XR environments.*

AI-Driven Multilingual Communication

Speech-to-text transcription utilizes the Whisper AI model for highly accurate multilingual voice recognition, demonstrated effectively in English, French, and Greek languages. This real-time service is crucial for inclusive participation, especially benefiting hearing-impaired learners through continuous captioning (Fig. 2).

```
(venv) andrew-daskalos@Mac ASR % python -u test.py
Reading metadata...: 1696it [00:00, 7038.75it/s]
The attention mask is not set and cannot be inferred from input because pad token is same as eos token. As a consequence, you may observe unexpected
behavior. Please pass your input's `attention_mask` to obtain reliable results.
ASR test on file el_test_0/common_voice_el_27443549.mp3. Expected: Γιατί την αλυσίδα μού την είχε δώσει εμένα
Actual: ['γιατί την αλυσίδα μου την είχε δώσει εμένα']
```

Figure 2: *Speech-to-text workflow demonstrating multilingual voice transcription accuracy using Whisper AI.*

```
 | > trim_db:60
 | > do_sound_norm:False
 | > do_amp_to_db_linear:True
 | > do_amp_to_db_mel:True
 | > do_rms_norm:False
 | > db_level:None
 | > stats_path:/root/.local/share/tts/vocoder_models--universal--libri-tts--fullband-melgan/scale_stats.npy
 | > base:10
 | > hop_length:256
 | > win_length:1024
 > Generator Model: fullband_melgan_generator
 > Discriminator Model: melgan_multiscale_discriminator
Select an option:
1: English text to English speech
2: English text to French speech
Enter your choice (1-2): 2
Enter text in English: Hello how are you?
Translated Text (French): Bonjour, comment allez-vous ?
 > Text splitted to sentences.
['Bonjour, comment allez-vous ?']
bɔ̃ ʒuʁ, kɔmɑ̃  ale vu ?
 [!] Character '̃' not found in the vocabulary. Discarding it.
 > interpolating tts model output.
 > before interpolation : (80, 75)
 > after interpolation : torch.Size([1, 80, 112])
 > Processing time: 0.6049902439117432
 > Real-time factor: 0.24384935264479773
Speech saved to output_french.wav
Applying noise reduction...
Processed audio saved to output_french_processed.wav
```

Figure 3: *Workflow illustrating integration of text-to-speech with language translation from English text to high-quality French audio.*

Text-to-speech functionality employs Tacotron2-DDC (Coqui TTS) paired with language translation (Helsinki-NLP opus-mt), providing high-quality, natural audio output in multiple languages. This approach significantly enhances multilingual accessibility and user engagement within immersive learning scenarios (Fig. 3).

Accessible Communication Through Sign-Language Translation

INFINITY addresses accessibility for deaf users by developing a text-to-sign translation pipeline that integrates AI-based animation with skeletal gesture tracking. We created a structured dataset using MediaPipe Holistic and OpenCV to extract pose and hand landmarks from International Sign Language (ISL) video samples. The extracted frame-level skeletal data, including 33 pose and 21 hand landmarks per hand, is serialized into JSON format and used to drive avatar animations within XR environments. This pipeline enables gesture-to-avatar translation with sub-300ms latency, ensuring smooth, natural sign delivery. The modular approach supports real-time integration and lays the groundwork for training ISL gesture recognition and generation models in immersive learning applications.

AI-Powered Session Summarization

Large Language Models (LLMs), specifically BART-Large CNN SamSum, enable effective session summarization by condensing complex XR interactions into concise, coherent narratives. This facilitates learning reinforcement and provides targeted feedback post-session, significantly enhancing the overall educational value of immersive XR activities (Fig. 4).

```
Input text: Number of words: 1274, Number of characters: 10818
Summary: Number of words: 147, Number of characters: 1090
Summary:
 In the past few decades, technological advancements have augmented the traditional methods of language education. Extended Reality (XR), enc
ompassing Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR), has significantly advanced foreign language learning. However
, gaps remain in engagement, accessibility, and inclusivity of users in XR-based language learning, as well as in pedagogical frameworks to g
uide the effective integration of XR technologies into language curricula. AI-driven 3D avatars interact with learners, providing personalize
d language instruction and facilitating conversational practice. AI for sign language recognition has been achieved, but the development of c
omprehensive systems remains constrained. Current XR technology integrates modular system with a focus on speech-to-text, text-to text, and s
peech and text-speech. It also provides flexibility and scalability across XR platforms. It can be used for deaf and hearing users. It's also
```

Figure 4: *Summarization workflow utilizing the BART-Large CNN SamSum model, demonstrating concise and coherent session summary generation.*

Immersive Educational Scenario

An immersive XR classroom (Fig. 5) demonstrates how INFINITY's AI-driven services converge seamlessly to create an inclusive educational environment. Here, 3D avatars employ multilingual speech recognition, text-to-sign translation, and multilingual narration to foster equitable, accessible learning interactions between deaf and hearing users.

Figure 5: *XR-based classroom featuring a 3D avatar delivering educational content, showcasing real-time multilingual translation, text-to-sign interpretation, and speech synthesis.*

Scalability Validation: A rigorous load-testing scenario simulating 1000 concurrent users validated platform scalability and robustness, maintaining response times below 800 milliseconds. This demonstrates INFINITY's capacity to support high-demand educational deployments reliably, leveraging cloud-native architectures and API-driven services.

Future Work and Studies

Future work on the INFINITY platform will focus on expanding and validating the AI services introduced, with a particular emphasis on refining sign language support and improving multilingual coverage. To address the current limitations in International Sign Language (ISL), we plan to significantly expand the gesture dataset by incorporating additional annotated video material and improving avatar expressiveness for more natural and grammatically accurate sign output. A series of controlled studies will be conducted with deaf and hearing learners to assess usability, accessibility, and learning outcomes using pre/post-tests, user satisfaction surveys, and cognitive load metrics. Ethical development remains a central priority. We are implementing data anonymization methods for voice and gesture inputs, enforcing secure API access, and applying fairness audits across AI components. Special attention will be paid to cultural and linguistic inclusivity in avatar design and sign language animation to avoid misrepresentation and ensure equitable learning experiences for all users

References

1. Shliakhtina, O., Kyselova, T., Mudra, S., Talalay, Y., & Oleksiienko, A. (2023). The effectiveness of the grammar translation method for learning english in higher education institutions. Eduweb. DOI:10.46502/issn.1856-7576/2023.17.03.12
2. Wu, H., Su, H., Yan, M., & Zhuang, Q. (2023). Perceptions of GrammarTranslation Method and Communicative Language Teaching Method Used in English Classrooms. Journal of English Language Teaching and Applied Linguistics. DOI:10.32996/jeltal.2023.5.2.12
3. Divekar*, R. et al. (2021). Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. Computer Assisted Language Learning, 35(9), 2332–2360. https://doi.org/10.1080/09588221.2021.1879162
4. Panagiotidis, Panagiotis. (2021). Virtual Reality Applications and Language Learning. International Journal for Cross-Disciplinary Subjects in Education. 12. 4447-4454. 10.20533/ijcdse.2042.6364.2021.0543.
5. Taborda, C.L., Nguyen, H., Bourdot, P. (2025). Engagement and Attention in XR for Learning: Literature Review. In: Reyes-Lecuona, A., et al. Virtual Reality and Mixed Reality. EuroXR 2024. Lecture Notes in Computer Science, vol 15445. Springer, Cham. https://doi.org/10.1007/978-3-031-78593-1_13

This research is supported by FSTP Funding from Horizon Europe under grant agreement no. 101070631 (UTTER). Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Directorate-General for Communications Networks, Content and Technology. Neither the European Union nor the granting authority can be held responsible for them.