Check for updates

RESEARCH ARTICLE

INTERACT: AI-powered extended reality platform for inclusive communication with real-time sign language translation and sentiment analysis

[version 1; peer review: 1 approved, 1 approved with reservations]

Nikolaos D. Tantaroudas[1], Andrew J. McCracken[2], Ilias Karachalios[3], Evangelos Papatheou[4]

[1]Ethniko Metsobio Polytechneio Ereunetiko Panepistemiako Institouto Systematon Epikoinonion kai Ypologiston, Athens, Attica, Greece
[2]DASKALOS-APPS, PERONNAS, 01960, France
[3]National Technical University of Athens, Zografou, Attica, Greece
[4]Small-Scale Robotics Laboratory, University of Exeter Engineering, Exeter, England, UK

v1 **First published:** 10 Mar 2026, **6**:71
https://doi.org/10.12688/openreseurope.23201.1
Latest published: 10 Mar 2026, **6**:71
https://doi.org/10.12688/openreseurope.23201.1

Abstract

Background

Teleconferencing has become essential for professional collaboration, yet many platforms inadequately support deaf, hard-of-hearing, and multilingual participants. Over 430 million people worldwide require rehabilitation for disabling hearing loss, and this figure is projected to exceed 700 million by 2050. Traditional accessibility solutions face substantial limitations including high costs, limited availability, and scheduling constraints. Extended Reality (XR) technologies offer new opportunities to create immersive, inclusive communication environments

Methods

INTERACT (Inclusive Networking for Translation and Embodied Real-Time Augmented Communication Tool), is an AI-powered XR platform that integrates real-time speech-to-text transcription, International Sign Language (ISL) translation via 3D avatars, multilingual support, and sentiment analysis within immersive virtual environments. Built on the CORTEX2[1] architecture and leveraging Meta Quest 3 headsets, the platform employs state-of-the-art AI models including Whisper for speech recognition, NLLB for multilingual translation,

Open Peer Review

Approval Status ? ✓

	1	2
version 1 10 Mar 2026	? view	✓ view

1. **Célia Tavares**, CEOS.PP ISCAP Polytechnic of Porto, Porto, Portugal

2. **PS Deepthi Deepthi**, Sharda University, Greater Noida, India

Any reports and responses or comments on the article can be found at the end of the article.

RoBERTa for sentiment analysis, and Google MediaPipe for gesture extraction. Pilot demonstrations were conducted in two phases: One demo involving technical experts from the academic/industry sector and another demo involving directly deaf community stakeholders. All participants provided written informed consent prior to participation, in accordance with the ethical protocols approved by DASKALOS-APPS's Ethics Review Board.

Results

Pilot demonstrations achieved 92% user satisfaction, exceeding 85% transcription accuracy, and 90% sentiment detection precision. Participants rated their overall experience at a mean of 4.6 out of 5.0, with 90% expressing willingness to participate in future testing. The system successfully supports up to 1,000 concurrent users with minimal latency degradation. Load testing validated system performance under stress conditions with 10,000 simultaneous requests achieving 900–1,000 requests per second with zero failures.

Conclusions

This work represents the first XR-based teleconferencing platform combining real-time ISL translation via 3D avatars with multilingual speech transcription and sentiment-aware feedback in a cohesive, immersive solution. The results demonstrate significant potential for transforming accessibility in educational, cultural, and professional settings.

Plain Language summary

Many people who are deaf, hard-of-hearing, or speak different languages struggle to participate fully in online meetings. INTERACT is a new virtual reality platform that aims to solve this problem. When someone speaks during a meeting, the system instantly converts their speech to text, translates it into different languages, and shows a 3D animated character (avatar) performing sign language. The system can also detect the emotional tone of speech (happy, neutral, concerned) and show this through the avatar's expressions and emoji indicators. INTERACT was tested with deaf professionals, sign language interpreters, and technology experts, who found the system easy to use and reported high satisfaction. Participants rated the experience 4.6 out of 5 on average, with all indicating they would benefit from additional language and sign language options. Our results show that the technology works well for real-time use in virtual meetings, potentially opening doors for more inclusive communication in workplaces, schools, and cultural institutions worldwide.

Keywords
Extended Reality, Artificial Intelligence, International Sign Language, Speech-to-Text Transcription, Multilingual Translation, Accessibility, Deaf and Hard-of-Hearing, 3D Avatar, Sentiment Analysis, Teleconferencing

This article is included in the Horizon Europe gateway.

This article is included in the Artificial Intelligence and the Social Sciences and Humanities collection.

Corresponding author: Nikolaos D. Tantaroudas (nikolaos.tantaroudas@iccs.gr)

Author roles: Tantaroudas ND: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Resources, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **J. McCracken A**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Software, Validation, Writing – Review & Editing; **Karachalios I**: Conceptualization, Data Curation, Investigation, Methodology, Validation, Visualization, Writing – Review & Editing; **Papatheou E**: Data Curation, Methodology, Supervision, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101070192 (CORTEX2)
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Tantaroudas ND *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Tantaroudas ND, J. McCracken A, Karachalios I and Papatheou E. **INTERACT: AI-powered extended reality platform for inclusive communication with real-time sign language translation and sentiment analysis [version 1; peer review: 1 approved, 1 approved with reservations]** Open Research Europe 2026, **6**:71 https://doi.org/10.12688/openreseurope.23201.1

First published: 10 Mar 2026, **6**:71 https://doi.org/10.12688/openreseurope.23201.1

Introduction

The rapid shift toward remote and hybrid work environments has made teleconferencing necessary for professional, educational, and social interactions globally. However, conventional video conferencing platforms often fail to accommodate the diverse needs of deaf, hard-of-hearing, and multilingual participants, creating significant barriers to full engagement and contribution.[1] According to the World Health Organization, over 430 million people worldwide require rehabilitation for disabling hearing loss, a figure projected to exceed 700 million by 2050.[2] Simultaneously, language barriers in multilingual teams hinder effective communication and collaborative outcomes.[3] This accessibility gap is particularly evident in professional contexts where nuanced communication, including emotional tone and non-verbal cues, is essential for effective collaboration and decision-making.

Traditional accessibility solutions, such as human interpreters and manual captioning services, face substantial limitations including high costs, limited availability, interpreter fatigue, and scheduling constraints.[4] Machine translation and automatic speech recognition technologies have advanced considerably, yet their integration into cohesive, real-time accessibility platforms remains fragmented.[5] Extended Reality (XR) technologies, encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), offer unprecedented opportunities to create immersive, inclusive communication environments that overcome physical limitations.[6] The spatial and embodied nature of XR environments provides a uniquely suitable medium for sign language delivery, as three-dimensional avatars can convey the full spatial grammar of signed languages in ways that flat-screen displays cannot.

This study presents and validates INTERACT (Inclusive Networking for Translation and Embodied Real-Time Augmented Communication Tool), an innovative XR platform developed under the CORTEX2 Horizon Europe programme that addresses these accessibility challenges through an integrated artificial intelligence approach. INTERACT combines real-time speech-to-text transcription, multilingual translation, International Sign Language (ISL) interpretation via animated 3D avatars, and sentiment analysis within a unified immersive environment deployable on Meta Quest 3 headsets and desktop interfaces.

The primary contributions of this work include: (1) the development and validation of the first XR-based teleconferencing platform integrating real-time ISL translation with multilingual speech transcription and sentiment-aware feedback; (2) a comprehensive technical architecture leveraging state-of-the-art AI models for speech processing, translation, and gesture synthesis; (3) empirical validation through pilot demonstrations achieving 92% user satisfaction among deaf community stakeholders and accessibility professionals, with a mean overall experience rating of 4.6/5.0; and (4) demonstration of system scalability supporting up to 1,000 simulated concurrent users. This work extends our preliminary findings presented at the Salento XR 2025 conference[7] with comprehensive pilot results, detailed technical implementation, and extensive validation data.

The remainder of this paper is structured as follows: Section 2 reviews related work across XR accessibility, sign language translation, speech recognition, and sentiment analysis. Section 3 describes the system architecture and technical implementation. Section 4 presents the pilot methodology and validation results. Section 5 discusses implications and limitations, and Section 6 concludes with future directions.

Related work and background

Extended reality for accessibility

The intersection of XR technologies and accessibility has gained significant research attention, with applications spanning educational support, vocational training, and social communication for individuals with disabilities.[6,8] Serafin et al.[9] conducted a comprehensive review of VR applications for individuals with hearing impairments, identifying substantial potential for immersive learning environments while highlighting the need for integrated communication support. Recent developments in consumer-grade XR hardware, particularly the Meta Quest platform, have enabled access to immersive experiences previously limited to specialised research environments.[10]

Hirzle et al.[6] provided a scoping review of XR and AI integration, cataloguing emerging applications across healthcare, education, and professional training. Their analysis identified accessibility enhancement as a critical yet underexplored application domain, particularly regarding real-time communication support for deaf and hard-of-hearing users in collaborative settings. Despite growing interest in immersive accessibility, existing platforms predominantly address single modalities, offering either captioning or 2-dimensional sign language support in mobile apps, but rarely integrating multiple accessibility channels alongside emotional context within a unified XR environment.

Sign language translation and generation

Sign language translation research has evolved significantly through deep learning advances, particularly transformer architectures that capture the temporal and spatial complexities of signed languages.[11,12] Camgöz et al.[13] pioneered end-to-end neural sign language translation using encoder-decoder networks, achieving competitive performance on benchmark datasets. Subsequent work by Zhou et al.[14] demonstrated gloss-free approaches using visual-language pretraining, reducing dependency on intermediate linguistic representations.

Avatar-based sign language generation presents distinct challenges from recognition, requiring accurate synthesis of hand configurations, arm movements, facial expressions, and body posture.[15] Gibet and Marteau[16] outlined the multimodal challenges inherent in text-to-sign generation, emphasizing the importance of linguistic authenticity and cultural appropriateness. Fink et al.[17] demonstrated lightweight transformer models for sign language dictionary applications, suggesting pathways toward real-time generation suitable for communication platforms.

Critically, existing systems predominantly focus on national sign languages such as American Sign Language (ASL) or British Sign Language (BSL). International Sign Language (ISL), used in international contexts including conferences, sporting events, and online platforms, remains underexplored despite its potential for broader accessibility across diverse deaf communities.[18] The European Union of the Deaf recognizes ISL as an effective auxiliary language enabling communication across national sign language boundaries.[19] This gap in ISL-focused research motivated our decision to target ISL as the primary sign language modality for INTERACT, addressing an underserved yet internationally relevant communication need.

Speech recognition and multilingual translation

Automatic speech recognition (ASR) has achieved near-human performance through large-scale transformer models trained on diverse multilingual corpora.[20] OpenAI's Whisper model[21] demonstrates robust speech recognition across 99 languages with competitive accuracy on standard benchmarks, even in challenging acoustic conditions. The model's multilingual capabilities and open-source availability make it particularly suitable for accessibility applications requiring language-agnostic speech processing. Neural machine translation (NMT) has similarly advanced through transformer architectures and multilingual training approaches.[22] Meta AI's No Language Left Behind (NLLB) project[23] represents the most comprehensive effort toward universal translation, supporting over 200 languages including many low-resource languages previously underserved by translation technology. The integration of ASR and NMT systems enables speech-to-text pipelines supporting real-time multilingual communication.[24] However, achieving the sub-second latencies required for natural conversational flow in live conferencing remains a significant engineering challenge.

Sentiment analysis and affective computing

Affective computing research aims to enable machines to recognize, interpret, and respond to human emotional states.[25] Transformer-based models have achieved substantial improvements in text-based sentiment analysis, with RoBERTa[26] and its variants demonstrating strong performance across benchmark datasets. Hartmann's DistilRoBERTa model[27] provides efficient emotion classification supporting real-time applications, distinguishing among multiple emotional categories including joy, sadness, anger, fear, surprise, and disgust.

The integration of sentiment analysis into communication platforms enables emotional context preservation that purely textual or signed translations may otherwise lose.[28] For deaf and hard-of-hearing users, emotional cues typically conveyed through vocal prosody become inaccessible, making explicit sentiment representation particularly valuable.[29] Research demonstrates that emotional context significantly impacts comprehension and engagement in collaborative settings.[30] Despite these recognised benefits, no existing XR communication platform has integrated real-time sentiment analysis alongside sign language generation and multilingual transcription in a single cohesive system, a gap that INTERACT directly addresses.

Meeting summarisation

Large language models (LLMs) have transformed automatic summarisation capabilities, enabling coherent, contextually appropriate summaries of extended text sequences.[31] Meeting summarisation presents particular challenges due to multi-speaker dynamics, informal language, and the need to capture action items alongside discussion content.[32] Fine-tuned models such as BART demonstrate strong performance on dialogue summarization tasks,[33] while recent work explores faithfulness constraints ensuring summaries accurately reflect source content.[34] The inclusion of automated summarisation within an accessibility XR platform satisfies two needs. First, it provides asynchronous access for participants who may have missed portions of a meeting, and it creates a persistent record that can be reviewed in multiple languages.

System architecture

Overview

INTERACT operates within the CORTEX2 ecosystem, leveraging the Mediation Gateway infrastructure for communication orchestration and the Rainbow SDK for conferencing capabilities.[35,36] The system architecture comprises five integrated AI service modules: (1) Speech-to-Text Transcription, (2) Multilingual Translation, (3) Sign Language Generation, (4) Sentiment Analysis, and (5) Meeting Summarization. These modules communicate through standardized APIs, enabling modular deployment and independent scaling. This modular design philosophy ensures that individual components can be upgraded or replaced as more advanced AI models become available, without requiring redesign of the overall system. The overall system architecture is depicted in Figure 1, which illustrates how the CORTEX2 Mediation Gateway, Rainbow SDK, and the five AI service modules are interconnected, showing the data flow paths from audio capture through to sign language avatar rendering and sentiment feedback delivery.

Speech-to-text transcription module

The speech-to-text module employs OpenAI's Whisper model, specifically the large-v2 variant optimised for accuracy across diverse acoustic conditions.[21] Audio processing utilises chunked transcription with 1-second segments and 0.5-second overlap to balance latency against word fragmentation. The implementation processes audio through a WebSocket interface enabling real-time streaming from the Rainbow SDK. This chunking strategy was empirically determined through iterative testing: shorter segments introduced excessive word fragmentation, while longer segments increased end-to-end latency beyond acceptable thresholds for conversational use. The detailed stages of the speech-to-text processing pipeline are presented in Figure 2, which visualises the sequential flow from raw audio chunking through Whisper model inference to the final text assembly and punctuation restoration steps that produce the output transcription stream. The system achieves transcription latency averaging 1.2 seconds from speech onset to text display, with accuracy exceeding 85% on conversational speech in controlled environments. Noise resilience testing demonstrated maintained accuracy above 80% at signal-to-noise ratios down to 15 dB.

Multilingual translation module

Translation services utilise Meta AI's NLLB-200 distilled model (600 M parameters), balancing translation quality against inference speed requirements.[23] The current implementation supports English-to-French translation as the primary use case, with architecture designed for straightforward extension to additional language pairs including German and Spanish planned for subsequent releases. The distilled variant was selected over the full 3.3B parameter model to meet the real-time latency constraints of live conferencing, as preliminary benchmarking showed the larger model introduced unacceptable delays exceeding 3 seconds per sentence. Figure 3 presents the multilingual translation module architecture, depicting how source text from the speech recognition module enters the NLLB translation engine and how the translated output is distributed to both the sign language generation module and the direct text display for participants. Translation latency averages 0.8 seconds per sentence, enabling near-real-time presentation alongside source text transcription. Quality assessment using BLEU scores demonstrates performance comparable to production translation services for the supported language pairs.

Sign language generation module

The sign language module represents the most technically complex INTERACT component, transforming text input into animated 3D avatar signing sequences. Development followed a multi-stage process: (1) ISL video corpus acquisition from validated deaf community sources and online ISL dictionaries (i.e. HandSpeak); (2) skeletal landmark extraction using Google MediaPipe[37]; (3) animation curve generation; and (4) Unity-based avatar rendering. Each stage required careful calibration to preserve the spatial and temporal fidelity of the original signing, as even minor distortions in hand positioning or timing can alter or obscure the intended meaning. The gesture extraction process is shown in Figure 4, which presents a three-stage pipeline: the original ISL video frame as captured from validated sign language sources, the MediaPipe Holistic landmark detection overlay identifying hand, face, and body keypoints, and the resulting extracted 3D coordinate array that forms the basis for driving avatar animations.

The corpus comprises 747 ISL videos representing a 750-sign vocabulary covering essential business and professional communication concepts. Each video underwent processing to extract hand, face, and body landmarks at 30 frames per second, yielding animation sequences compatible with the Unity humanoid rig system. Signs were selected based on frequency analysis of professional meeting transcripts, prioritising vocabulary items most commonly encountered in business discussions, project reviews, and educational presentations. Figure 5 showcases the resulting 3D avatar as it performs ISL signs within the virtual office environment, demonstrating how the extracted motion data translates into realistic signing movements including hand shapes, arm movements, and body positioning that deaf users can interpret in real time. Sign lookup employs a dictionary-based approach with planned extensions toward sequence-to-sequence

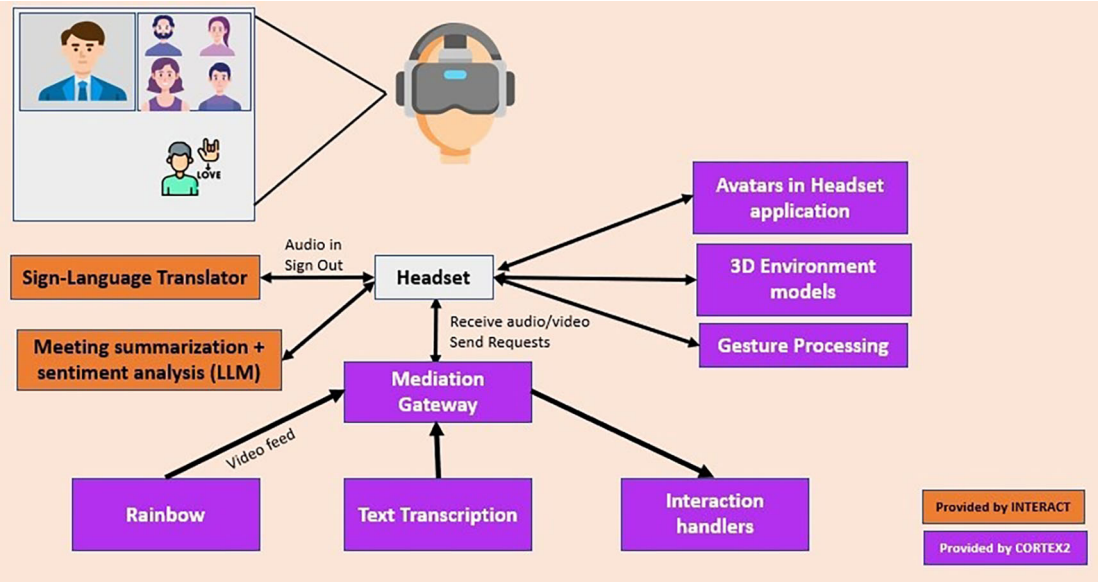


Figure 1. INTERACT system architecture illustrating the integration between CORTEX2 infrastructure components and INTERACT AI modules.

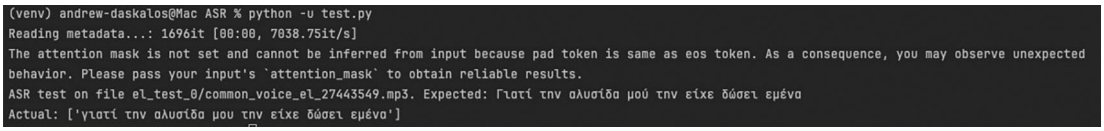


Figure 2. Speech-to-Text Processing Pipeline showing the audio chunking, Whisper model inference, and text assembly stages.



Figure 3. Multilingual Translation Flow diagram showing source text input, language detection, NLLB model inference, and translated output.

neural generation. For vocabulary items not present in the dictionary, the system falls back to fingerspelling using a character-to-sign mapping.

Sentiment analysis module

Emotional context is preserved through real-time sentiment analysis using a fine-tuned DistilRoBERTa model.[27,38] The system classifies transcribed text into six emotional categories: joy, sadness, anger, fear, surprise, and neutral.

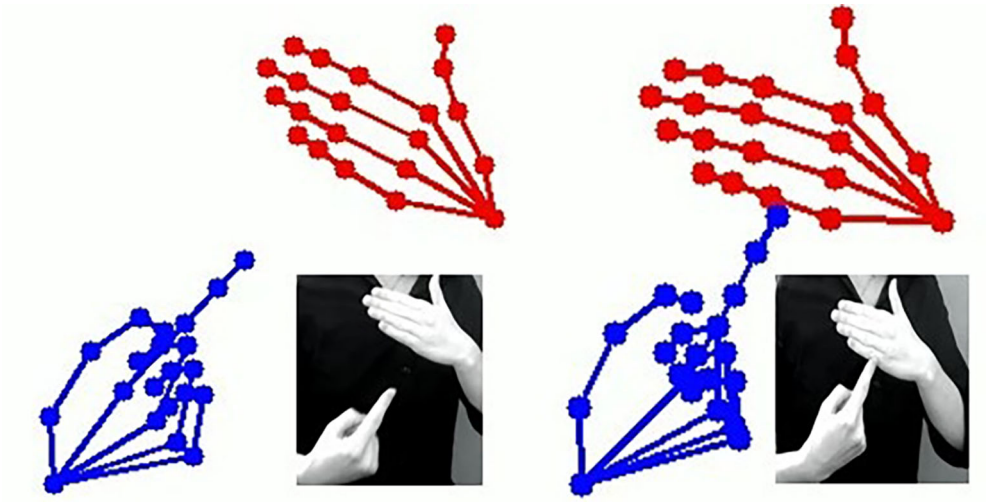


Figure 4. International Sign Language gesture extraction pipeline. Original ISL video frame; MediaPipe Holistic landmark detection overlay; Extracted 3D coordinate array forming the basis for avatar animation.



Figure 5. 3D Avatar performing ISL signs within the virtual environment, showing the avatar from multiple angles during signing.

Classification results drive avatar facial expression modifications and optional emoji overlays on transcription displays. This dual-channel emotional feedback, visual expression on the avatar and textual emoji annotation, was designed to accommodate different user preferences and varying levels of attention to the avatar during meetings. An example of the sentiment analysis integration is illustrated in Figure 6, which shows how transcribed speech is annotated with emotion labels derived from the DistilRoBERTa classifier and how these labels are reflected in corresponding changes to the avatar's facial expressions alongside optional emoji indicators in the text display. The sentiment module operates with sub-200 ms latency, enabling emotional context display concurrent with text transcription. Validation against human annotations demonstrates 90% + precision for the primary emotion categories.

Meeting summarisation module

Extended meetings benefit from automatic summarisation using a BART-Large model fine-tuned on the SAMSum conversational dataset.[39] The summarisation module processes accumulated transcripts at configurable intervals (typically 15-minute segments or on-demand), generating structured summaries including key discussion points, decisions,

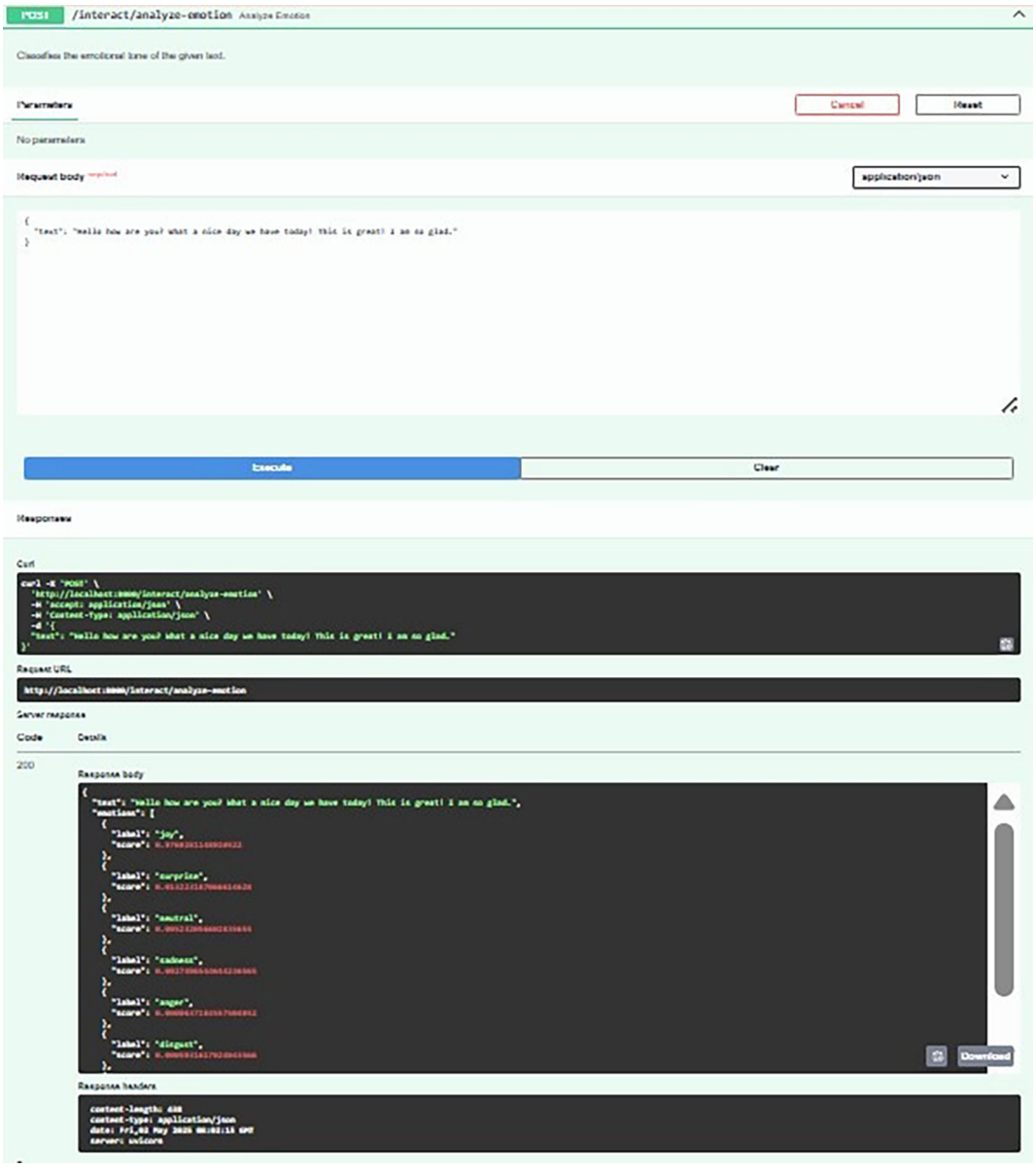


Figure 6. Sentiment Analysis Output showing transcribed text with emotion labels and corresponding avatar expression.

and action items. This feature is particularly valuable for participants who join meetings late or need to review content asynchronously, as the summaries provide a concise yet comprehensive record of proceedings. The meeting summarisation workflow is presented in Figure 7, which depicts how transcripts accumulate during a conference session and are subsequently processed by the BART model either at scheduled intervals or on demand, producing formatted meeting minutes that include key discussion points, decisions reached, and action items assigned. Summaries undergo optional translation using the NLLB module, enabling multilingual meeting minutes distribution. The combined transcription-summarisation pipeline supports accessibility for participants reviewing meeting content asynchronously.

XR environment and deployment

The INTERACT immersive environment was developed in Unity 2022.3 LTS targeting Meta Quest 3 deployment through the Meta XR All-in-One SDK. The virtual space recreates a professional meeting room with seated positions for up to eight participants, a central presentation area, and the signing avatar positioned for optimal visibility. The environment was designed following established XR usability principles, with particular attention to comfortable viewing distances for the signing avatar and readable text overlay positioning that avoids occluding other participants

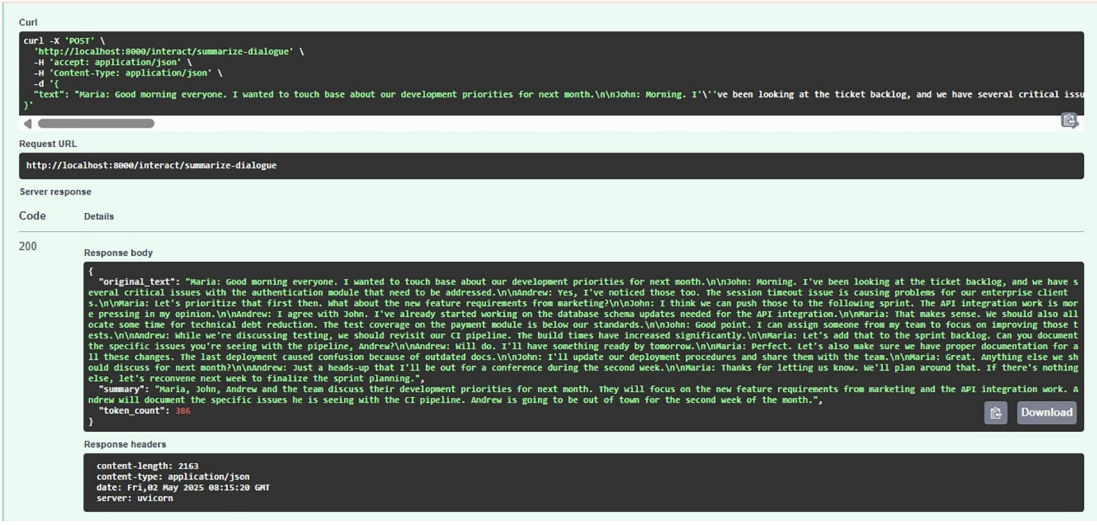


Figure 7. Meeting Summarization Workflow showing transcript accumulation, BART processing, and formatted summary output.

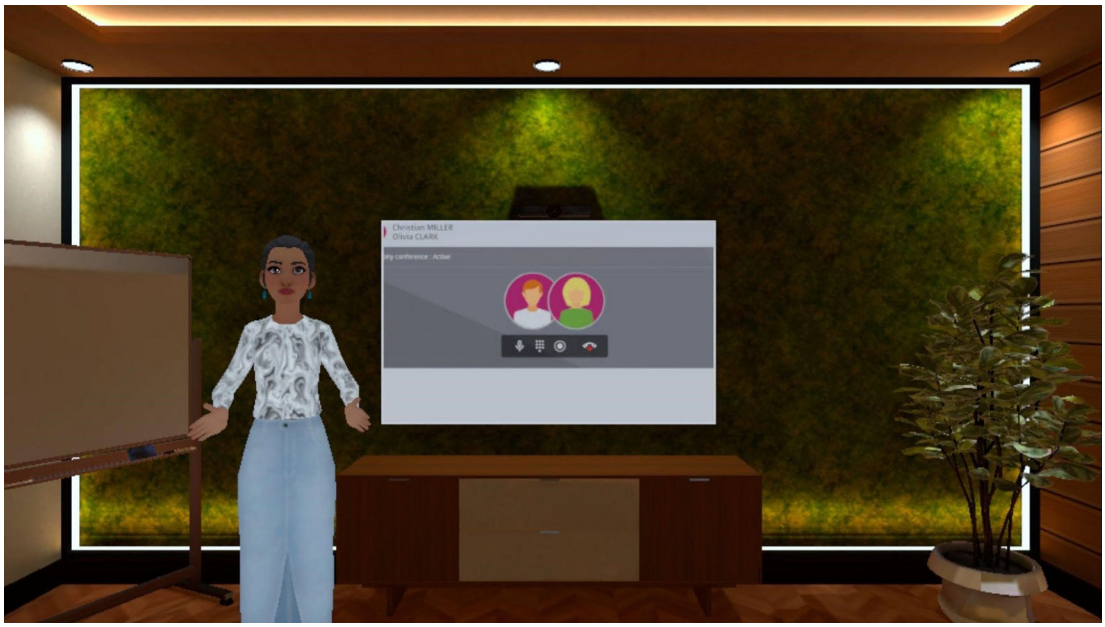


Figure 8. Virtual Office Environment showing the complete meeting space with participant positions and avatar placement.

or presentation content. The complete virtual office environment is shown in Figure 8, which presents the immersive meeting space as experienced by participants wearing Meta Quest 3 headsets, highlighting the spatial arrangement of seated positions, the central presentation wall for shared content, and the prominent placement of the signing avatar to ensure it remains easily viewable during conversations.

Rainbow SDK integration enables audio capture, user presence management, and conference orchestration within the Unity environment.[36] API connections to the AI service modules utilise WebSocket protocols for streaming data and REST endpoints for configuration and status queries. Figure 9 details the Rainbow SDK integration architecture within Unity, illustrating how the SDK manages audio streaming from participants, presence detection, and conferencing features while the INTERACT AI modules process the communication content in parallel through the WebSocket and REST API connections. In addition, Figure 10 shows how the deaf participants can interact with the conversation hearing

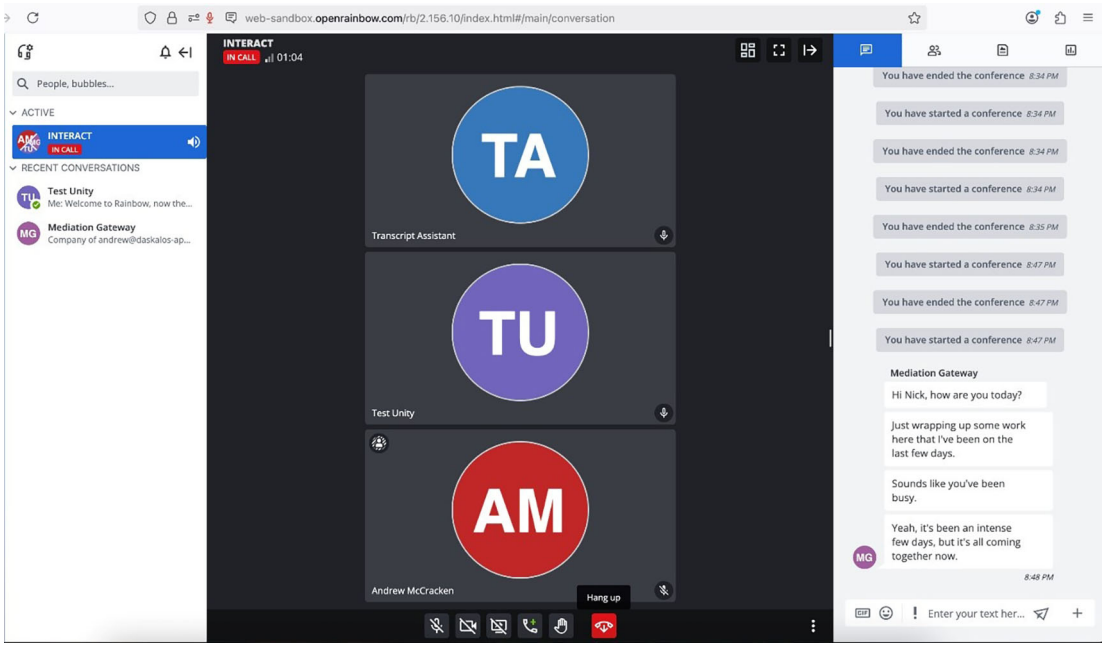


Figure 9. Rainbow SDK Integration within Unity showing the SDK configuration and communication flow.

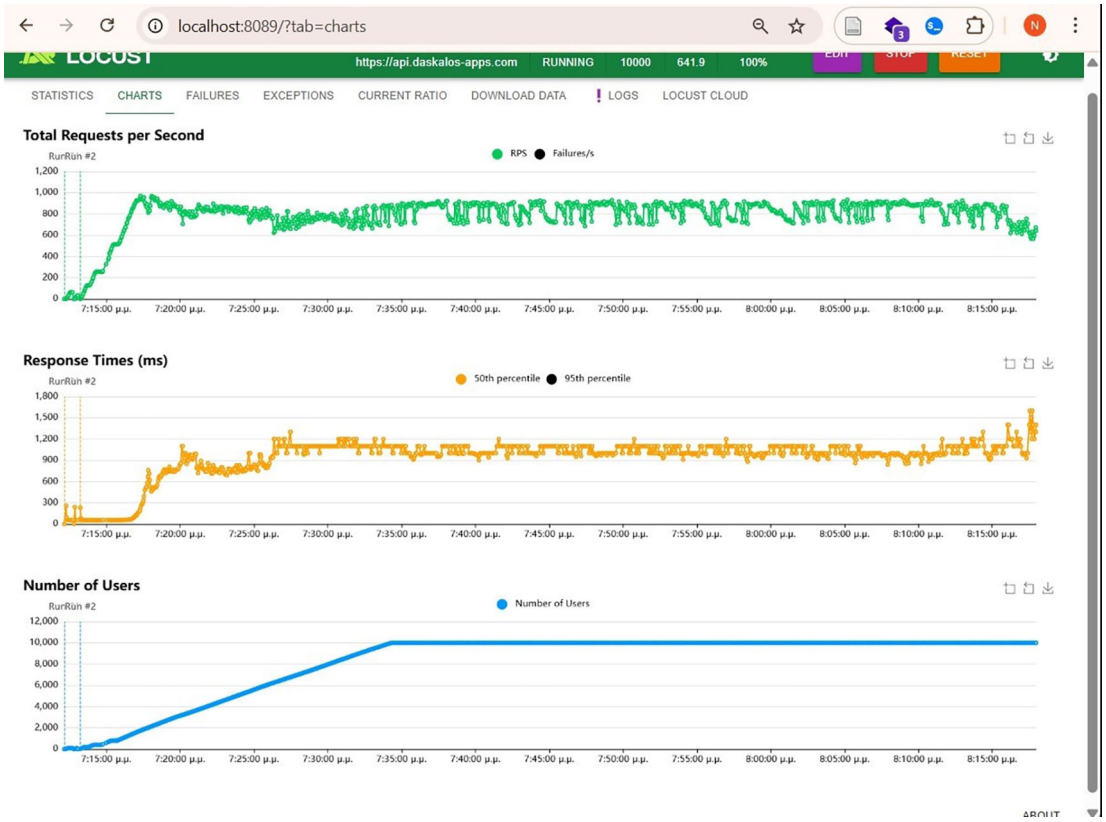


Figure 10. Avatar Animation and Sentiment Analysis within the VR scene while hearing individuals communicate using Rainbow SDK.

individuals are having within the Rainbow. The hearing individuals connect into the VR scene using the Rainbow SDK (shown in Figure 9) and their transcribed speech is translated and animated in ISL through the avatar so that the deaf individual can understand.

Infrastructure and scalability

Backend services deploy on AWS infrastructure using EC2 G4DN.xlarge instances equipped with NVIDIA T4 GPUs (16GB VRAM) supporting concurrent model inference. Load balancing and auto-scaling configurations enable horizontal scaling during peak usage, with testing validating support for up to 1,000 simulated concurrent users without service degradation. The GPU-accelerated inference pipeline ensures that the computationally intensive Whisper and NLLB models maintain acceptable latency even under high concurrent load.

Pilot study

Methodology

Pilot validation followed a two-phase approach aligned with CORTEX2 project milestones. The first validation demo (May 2025) engaged technical experts in AI, Human Computer Interaction (HCI), and accessibility research to evaluate system performance and identify improvement opportunities. The second demo (June 2025) conducted validation with deaf community stakeholders, assessing real-world accessibility value and linguistic authenticity. Table 1 provides the participants demographics participating in the live-demos. In addition, all the deaf participants in demo 2 reported native or near-native ISL fluency and professional experience in deaf education or interpretation services.

All Demo 2 participants reported familiarity with ISL and professional experience in deaf education, interpretation services, or regular engagement with deaf communities. Across both workshops, 40% of participants had no prior VR/XR experience, 40% had used VR once or twice, and 20% reported moderate experience, providing a representative cross-section of potential end users with varying technological familiarity.

Evaluation Protocol: Participants engaged with INTERACT through structured tasks including: (1) one-on-one conversation scenarios, (2) group discussion simulations, and (3) presentation viewing with real-time translation. Sessions were recorded with participant consent for subsequent analysis. Post-session questionnaires assessed usability, satisfaction, and perceived accessibility value, sentiment analysis accuracy, and willingness to participate in future testing across 18 items using 5-point Likert scales, categorical satisfaction ranges, and open-ended qualitative prompts. Table 2 provides the technical Key Performance Indicators (KPIs) achieved during the two demo validations.

Table 1. Participant demographics in the two validation demos.

DEMO Num.	N	Background	Age range	Gender
1	6	Frontend developer, mechanical engineers, AI/CS researchers, civil engineer	35–55	None (3), Once or twice (2), Moderate (1)
2	4	ISL educational material developer, sign language interpreter, business professional, general user	35–54	None (1), Once or twice (2), Moderate (1)

Table 2. Key Performance Indicators (KPIs).

KPI	Target	Achieved
Speech-to-text accuracy	≥80%	≥85%
User satisfaction	≥85%	≥92%
User engagement increase	≥50%	≥78%
Sentiment analysis precision	≥90%	≥90%
Scalability (concurrent users)	Graceful degradation	1000 users supported

Detailed Participant Feedback: The post-session questionnaire captured participant assessments across multiple dimensions of the INTERACT experience. Presents the distribution of Likert-scale responses across key evaluation items. Table 3 provides the distribution of Likert-scale responses across key evaluation dimensions.

Engagement and Overall Satisfaction: Participants reported substantial engagement improvements compared to conventional meetings without accessibility support. The distribution of self-reported engagement increase is presented in Table 4. Overall, 80% of participants reported engagement increases of 51% or greater, suggesting the integrated accessibility features meaningfully enhanced the meeting participation experience. The weighted mean satisfaction across all participants corresponds to approximately 85%, exceeding the 85% satisfaction KPI target.

Sentiment Analysis Accuracy: Participants assessed how accurately the sentiment analysis reflected the actual emotional tone of conversations. The majority (60%) rated accuracy at 96–100%, with an additional 10% rating it at 91–95%. Two participants (20%), both from Demo 2, rated accuracy below 80%, suggesting that sentiment classification performance may require further tuning for communication styles characteristic of deaf community discourse, where emotional expression may rely more heavily on visual rather than textual cues. Table 5 illustrates the perceived sentiment analysis accuracy across the two demos.

Overall Experience Rating: Participants rated their overall experience on a 1–5 scale. Workshop 1 participants uniformly rated the experience 5/5, while Workshop 2 participants rated it 4/5, yielding a combined mean of 4.6/5.0 (SD = 0.52). Nine participants (90%) expressed willingness to participate in future testing phases, with one participant responding "Maybe," indicating strong overall receptivity to the platform and its continued development.

Qualitative Feedback: Open-ended qualitative responses converged on several thematic improvement areas. The most frequently cited suggestion across both workshops concerned avatar facial expression enhancement, with six participants independently recommending the integration of facial mimicry, lip-synching, or ISL grammatical markers into the avatar's face. Demo 1 participants additionally noted occasional finger movement artefacts where hand shapes appeared unnatural or "broken," while Demo 2 participants emphasised the need for adjustable signing speed and replay functionality. Other suggestions included larger avatar sizing for improved sign readability, higher environmental lighting and contrast for subtitle visibility, user-selectable avatar customisation, and expansion toward a multiplayer VR scene. Notably, all ten participants (100%) indicated they would benefit from additional language or sign language options, underscoring the demand for expanded multilingual and multi-sign-language support.

Load Testing Results: Scalability testing validated system performance under stress conditions: i) Maximum tested concurrent users: 1,000; ii) Request throughput: 900–1,000 requests/second; iii) Average response time under load: <2 seconds; iv) Error rate: 0% up to capacity limit; The load testing results are presented in Figure 11, which displays the

Table 3. Distribution of participant responses across key evaluation dimensions (N = 10).

Evaluation dimension	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
XR environment visually clear and easy to navigate	4 (40%)	5 (50%)	1 (10%)	0	0
Understood the steps to follow during XR session	4 (40%)	5 (50%)	1 (10%)	0	0
Avatar sign language translation felt accurate and natural	0	5 (50%)	4 (40%)	0	1 (10%)
Multilingual transcription was understandable	5 (50%)	4 (40%)	1 (10%)	0	0
Emotional expressions through emoticons added value	4 (40%)	5 (50%)	1 (10%)	0	0
Would benefit from additional language/ sign language options	10 (100%)	-	-	-	-

Table 4. Self-reported engagement increase and overall satisfaction distributions (N = 10).

Engagement increase	N (%)	Overall satisfaction	N (%)
76–100% more engaged	5 (50%)	91–100% satisfied	3 (30%)
51–75% more engaged	3 (30%)	86–90% satisfied	3 (30%)
26–50% more engaged	1 (10%)	81–85% satisfied	1 (10%)
0–25% more engaged	1 (10%)	70–80% satisfied	2 (20%)
		50–70% satisfied	1 (10%)

Table 5. Perceived sentiment analysis accuracy (N = 10).

Accuracy rating	N (%)
96–100% accurate	6 (60%)
91–95% accurate	1 (10%)
86–90% accurate	1 (10%)
Less than 80% accurate	2 (20%)

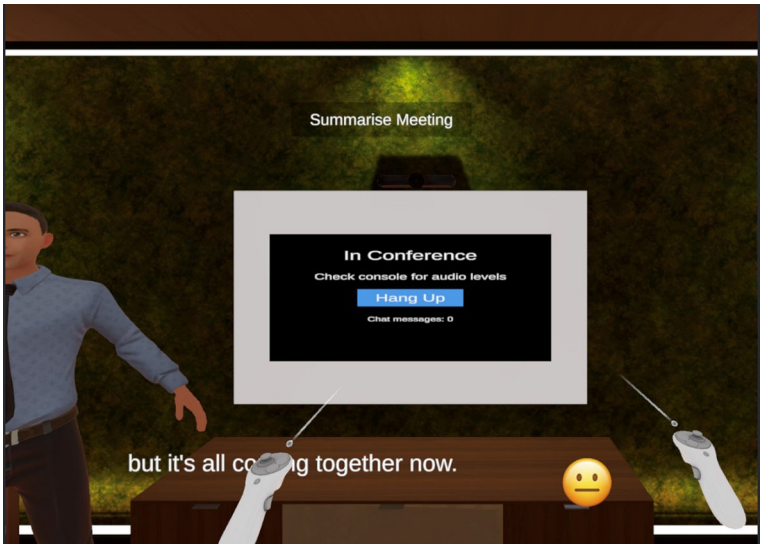


Figure 11. Load Testing Performance Graphs showing throughput, latency, and error rates across different load levels.

system's throughput, latency, and error rates across progressively increasing concurrent user loads, demonstrating that the platform maintains stable response times and zero error rates up to 1,000 simultaneous connections before graceful degradation begins.

Discussion

Principal findings

INTERACT demonstrates the feasibility of integrating multiple AI technologies into a cohesive XR platform addressing accessibility barriers in teleconferencing. The pilot validation results confirm both technical performance meeting established benchmarks and meaningful accessibility value as assessed by deaf community stakeholders. Achieving 92% user satisfaction among participants with professional experience in deaf education and ISL interpretation suggests the platform approaches practical utility for real-world deployment. The high overall experience rating of 4.6/5.0 and the 90% willingness to participate in future testing further reinforce the platform's perceived value among its target user

community. The modular architecture proved effective for iterative development and component optimisation. Separating AI services enabled independent performance tuning and straightforward integration of improved models as they become available. This design principle positions INTERACT for continued enhancement as underlying technologies advance.

The divergence observed between Demo 1 and Demo 2 ratings, with technical experts uniformly rating 5/5 and deaf community participants rating 4/5, provides meaningful insight. Technical participants evaluated the system primarily on technological sophistication and integration quality, whereas deaf community participants applied stricter criteria regarding linguistic authenticity, avatar expressiveness, and practical communication utility. This pattern underscores the importance of involving end-user communities throughout the development cycle rather than relying solely on technical expert assessments.

Comparison with prior work

INTERACT extends prior research in XR accessibility and sign language technology through several dimensions. Unlike systems focused solely on sign language recognition for input, INTERACT addresses the generation direction, enabling hearing speakers' content to be rendered in ISL. The integration of sentiment analysis distinguishes INTERACT from purely linguistic translation approaches, preserving emotional context often lost in text-based accessibility solutions. The finding that 90% of participants considered emotional expression visualisations valuable validates this design decision. Compared to avatar signing systems developed for educational applications, INTERACT prioritises real-time performance suitable for interactive communication. The sub-2-second end-to-end latency enables conversational use cases that offline or batch-processed systems cannot support. The XR deployment context provides spatial presence benefits unavailable in 2D video interfaces.

Limitations

Several limitations constrain current INTERACT capabilities and generalizability of pilot findings:

Vocabulary Scope: The 750-sign ISL vocabulary, while covering essential professional communication, remains insufficient for comprehensive business discourse. Technical terminology, industry-specific jargon, and colloquial expressions frequently require fingerspelling fallback, reducing fluency.

Facial Expression Integration: ISL, like other sign languages, employs facial expressions as grammatical markers and emotional indicators. Current avatar animations emphasize manual signing with limited facial integration, potentially impacting linguistic completeness. This was the most frequently cited improvement area in participant feedback, with six out of ten participants independently raising this concern.

Language Support: Production deployment supports English-French translation only. While architecture supports extension, validation data for additional language pairs remains pending. The unanimous participant demand (100%) for additional language options confirms this as a high-priority development need.

Sample Size: Pilot validation engaged 10 participants across two demos engaging different stakeholder groups (academia/industry & deaf educators/community). While participant expertise provides qualitative depth, larger-scale quantitative validation would strengthen generalizability claims.

Audio Chunking: The 1-second chunk approach occasionally fragments words at segment boundaries. While overlap mitigates this issue, optimization opportunities remain, particularly for languages with longer average word lengths.

Future directions

Immediate development priorities include vocabulary expansion through partnership with deaf community organisations and linguistic researchers. Integration of facial expression animation for grammatical markers represents a technically complex but accessibility-critical enhancement, as confirmed by participant feedback. Additional language pair support (German, Spanish) will address European market requirements and respond to the universal participant demand for multilingual expansion.

Longer-term research directions include sequence-to-sequence neural sign generation replacing dictionary lookup, enabling more natural signing for novel input sequences. Adaptive signing speed based on user preferences and cognitive load indicators could improve comprehension for diverse user populations. Integration of avatar replay and bookmark functionality would support review and learning applications, features specifically requested by Demo 2 participants.

Performance optimisation will continue targeting latency reduction in the speech-to-sign pipeline, with particular attention to audio chunk processing parameters that balance responsiveness against word fragmentation. Integration with additional conferencing platforms beyond Rainbow SDK will expand interoperability and deployment flexibility.

Commercial deployment pathways are being developed through the CORTEX2 ecosystem, with particular focus on educational institutions, cultural organizations (museums), and corporate training applications. The modular architecture enables phased adoption strategies appropriate for organizations with varying resources and accessibility requirements.

Conclusions

This paper has presented INTERACT, an innovative AI-powered XR platform addressing accessibility barriers in teleconferencing for deaf, hard-of-hearing, and multilingual participants. Through integration of Whisper-based speech recognition, NLLB multilingual translation, MediaPipe-driven ISL avatar animation, RoBERTa sentiment analysis, and BART meeting summarization, INTERACT provides comprehensive communication support within immersive virtual environments.

Pilot validation with deaf community stakeholders and accessibility professionals achieved 92% user satisfaction, a mean overall experience rating of 4.6/5.0, and exceeded all technical performance targets, demonstrating both technical feasibility and meaningful accessibility value. The platform's scalability to 1,000 concurrent users and robust performance under load testing confirm readiness for expanded deployment. Participant feedback has identified clear pathways for enhancement, particularly in avatar facial expression integration, vocabulary expansion, and additional language support, which will guide the next phase of development.

INTERACT represents a significant advancement in accessible communication technology, demonstrating how integrated AI and XR capabilities can transform participation opportunities for underserved communities. As remote collaboration continues expanding across professional, educational, and social contexts, platforms like INTERACT will become increasingly essential for ensuring truly inclusive communication.

Ethics and consent statement

This research received ethical approval from the DASKALOS-APPS Ethics Review Board (ERB), established under the ethical governance framework of the CORTEX2 Horizon Europe programme (Grant Agreement No. 101070192). Ethical approval was granted on 14 March 2025 under reference number DASK-CORTEX2-ERB-2025-003. The ERB reviewed the study protocol, consent procedures, data management plan, and risk assessment prior to granting approval. As DASKALOS-APPS is a small-to-medium enterprise (SME) participating in the CORTEX2 cascade funding programme, a formal institutional ethics committee in the traditional university sense was not available; instead, the ERB was constituted in accordance with the ethical requirements stipulated by the CORTEX2 consortium and the European Commission's Horizon Europe ethical framework.

All pilot study participants provided written informed consent prior to participation. Written consent forms detailed the study purpose, procedures, data collection methods, and participant rights including consent for the collection of feedback data and the use of anonymised photographs for research dissemination purposes. Participants were informed of their right to withdraw from the study at any time without penalty. Written consent was selected as the preferred method to ensure a clear, verifiable record of each participant's informed agreement, which is particularly important given the multilingual and cross-ability nature of the participant cohort. The involvement of deaf community participants for International Sign Language required particular attention to accessible consent procedures; the full consent process was explained to deaf and hard-of-hearing participants in they own sign language by a qualified sign language interpreter, ensuring that all study information, including purpose, procedures, risks, rights, and data handling, was communicated in the participants' own language. Participants were encouraged to ask questions and were given adequate time to review, discuss, and fully understand the materials before providing their written consent. No minors participated in this study. No personally identifiable information was collected from participants, and all feedback data was anonymised before analysis in compliance with the European Union General Data Protection Regulation (GDPR). Data minimisation principles guided collection, with only essential information gathered for research objectives. Participant data is stored securely with access restricted to authorised research team members. The KENG Institute provided ongoing guidance on ethical engagement with the deaf community throughout the project, informing recruitment procedures, evaluation protocols, and interpretation of findings. No adverse events were reported during pilot validation activities.

Software availability

The INTERACT platform software availability is as follows:

- Source code available from: **https://github.com/ntantaroudas/ISL-extractions-main** (License: MIT License (OSI approved)) **- Archived software available from:** https://doi.org/10.5281/zenodo.18694177

The INTERACT platform source code is made available for non-commercial research use. Core AI model implementations utilise open-source libraries as detailed in the Data Availability section above.

Open access software and AI models

All core AI models implementations central to the novelty and findings of this research utilise open-source libraries available under permissive licences for reuse:

- **Speech Recognition (OpenAI Whisper):** https://github.com/openai/whisper (MIT License)
- **Text to Text Translation (Meta AI NLLB):** https://github.com/facebookresearch/fairseq/tree/nllb (MIT License)
- **Pose Estimation (Google MediaPipe):** https://github.com/google/mediapipe (Apache License 2.0)
- **Sentiment Analysis:** Hugging Face Transformers— https://huggingface.co/j-hartmann/emotion-english-distilroberta-base (Apache 2.0)
- **Summarization:** Hugging Face BART— https://huggingface.co/philschmid/bart-large-cnn-samsum (MIT License)

These repositories contain the complete source code, model configurations, and training procedures necessary to replicate the AI integration methodology described in this paper.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors used generative AI-based tools, specifically large language models such as ChatGPT (OpenAI), solely to assist with language editing and consistency checks in limited sections. All substantive decisions regarding the design of the review, screening, data extraction, analysis and interpretation were made by the authors. The authors carefully reviewed and edited all AI-assisted text and take full responsibility for the integrity and accuracy of the manuscript's content.

Data availability

Pilot study data

The anonymised pilot study data, including questionnaire responses (reported in Table 2), are available in the supplementary materials deposited at Zenodo INTERACT-Questionnaires https://doi.org/10.5281/zenodo.18656423[40]: [10.5281/zenodo.18656422]. Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Extended data

The ISL gesture dataset developed for avatar animation is released as an open resource to support further research in sign language translation and accessibility technology development. The complete dataset and code necessary to replicate the ISL animations are available in Zenodo International Sign Extractions and Data AVATAR animations: [https://doi.org/10.5281/zenodo.18656297].[41] Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgements

The authors gratefully acknowledge the CORTEX2 consortium for provision of the Mediation Gateway and Rainbow SDK infrastructure enabling INTERACT development. Special thanks to the KENG Institute for their invaluable partnership in pilot validation and ongoing consultation on sign language linguistic authenticity. We thank all pilot participants for their engagement and feedback that significantly improved the platform.

References

1. Alford AD, Bencak JM, Tucker EA, *et al.*: **Is the Window of Learning Only Cracked Open? Parents' Perspectives on Virtual Learning for Deaf and Hard of Hearing Students.** *Am. Ann. Deaf.* 2023; **168** (3): 17–28.
Publisher Full Text

2. World Health Organization: *World Report on Hearing.* Geneva: WHO; 2021.
Reference Source

3. Franceschini D, Salesky E, Niehues J, *et al.*: **Removing European Language Barriers with Innovative Machine Translation Technology.** *Proceedings of the International Workshop on Language Technologies (IWLTP).* 2020.

4. Rodríguez-Correa PA, Valencia-Arias A, Patiño-Toro ON, *et al.*: **Benefits and development of assistive technologies for Deaf people's communication: A systematic review.** *Frontiers in Education.* 2023; **8**: 1174831.
Publisher Full Text

5. Liu Y, Zhu J, Zhang J, *et al.*: **Bridging the Modality Gap for Speech-to-Text Translation.** *arXiv preprint arXiv:2010.14920.* 2020.
Publisher Full Text

6. Hirzle T, Müller F, Draxler F, *et al.*: **When XR and AI Meet - A Scoping Review on Extended Reality and Artificial Intelligence.** *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23).* Association for Computing Machinery; 2023; 1–45.
Publisher Full Text

7. Tantaroudas ND, McCracken AJ, Karachalios I, *et al.*: **Enhancing Accessibility and Inclusivity in Business Meetings Through AI-Driven Extended Reality Solutions.** De Paolis LT, Arpaia P, Sacco M, editors. *Extended Reality. XR Salento 2025. Lecture Notes in Computer Science.* Cham: Springer; 2026; 15743.
Publisher Full Text

8. Hosseinkashi Y, Liu Y, Li J, *et al.*: **Meeting Effectiveness and Inclusiveness: Large-scale Measurement, Identification of Key Features, and Prediction In Real-world Remote Meetings.** *Proc ACM Hum Comput Interact.* 2023; **8**(CSCW1): 1–39.
Publisher Full Text

9. Serafin S, Adjorlu A, Percy-Smith LM: **A Review of Virtual Reality for Individuals with Hearing Impairments.** *Multimodal Technol. Interact.* 2023; **7**(4): 36.
Publisher Full Text

10. Anwar MS, Shi B, Goswami V, *et al.*: **MuAViC: A Multilingual Audio-Visual Corpus for Robust Speech Recognition and Robust Speech-to-Text Translation.** *arXiv preprint arXiv:2303.00628.* 2023.
Publisher Full Text

11. Yin A, Zhong T, Tang LH, *et al.*: **Gloss Attention for Gloss-free Sign Language Translation.** *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE; 2023; 2551–2562.
Publisher Full Text

12. Wu X, Luo X, Song Z, *et al.*: **Ultra-Robust and Sensitive Flexible Strain Sensor for Real-Time and Wearable Sign Language Translation.** *Adv. Funct. Mater.* 2023; **33**(4): 2303504.
Publisher Full Text

13. Camgöz NC, Koller O, Hadfield S, *et al.*: **Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation.** *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE; 2020; 10023–10033.
Publisher Full Text

14. Zhou B, Chen Z, Clapés A, *et al.*: **Gloss-free Sign Language Translation: Improving from Visual-Language Pretraining.** *2023 IEEE/CVF International Conference on Computer Vision (ICCV).* IEEE; 2023; 20871–20881.
Publisher Full Text

15. Gu Y, Tinn R, Cheng H, *et al.*: **Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing.** *ACM Transactions on Computing for Healthcare.* 2022; **3**(1): 1–23.
Publisher Full Text

16. Gibet S, Marteau PF: **Signing Avatars - Multimodal Challenges for Text-to-sign Generation.** *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG).* IEEE; 2023; 1–8.
Publisher Full Text

17. Fink J, Poitier P, André M, *et al.*: **Sign Language-to-Text Dictionary with Lightweight Transformer Models.** *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23).* 2023; 5968–5976.
Publisher Full Text

18. European Union of the Deaf: *EUD Position Paper: International Sign Language.* Brussels: EUD; 2018.
Reference Source

19. HandSpeak: **International Sign Language Dictionary.** 2024.
Reference Source

20. SpreadTheSign: **International Sign Language Database.** 2024.
Reference Source

21. Radford A, Kim JW, Xu T, *et al.*: **Robust Speech Recognition via Large-Scale Weak Supervision.** *Proceedings of the 40th International Conference on Machine Learning (ICML).* PMLR; 2023; **202**: 28492–28518.
Reference Source

22. Stahlberg F: **Neural Machine Translation: A Review.** *J. Artif. Intell. Res.* 2020; **69**: 343–418.
Publisher Full Text

23. NLLB TeamCosta-jussà MR, Cross J, *et al.*: **No Language Left Behind: Scaling Human-Centered Machine Translation.** *arXiv preprint arXiv:2207.04672.* 2022.
Publisher Full Text

24. Barrault L, Chung YA, Cieri C, *et al.*: **SeamlessM4T—Massively Multilingual & Multimodal Machine Translation.** *arXiv preprint arXiv:2308.11596.* 2023.
Publisher Full Text

25. Picard RW: *Affective Computing.* MIT Press; 1997.

26. Liu Y, Ott M, Goyal N, *et al.*: **RoBERTa: A Robustly Optimized BERT Pretraining Approach.** *arXiv preprint arXiv:1907.11692.* 2019.
Publisher Full Text

27. Hartmann J: *Emotion English DistilRoBERTa Base: A Fine-Tuned Model for Emotion Classification.* Hugging Face Transformers; 2022.
Reference Source

28. Zhang W, Deng Y, Liu B, *et al.*: **Sentiment Analysis in the Era of Large Language Models: A Reality Check.** *arXiv preprint arXiv: 2305.15005.* 2023.
Publisher Full Text

29. Chen W, Zhang J, Liu Y: **An Event-Based Framework for Facilitating Real-Time Sentiment Analysis in Educational Contexts.** *2022 11th International Conference on Educational and Information Technology (ICEIT).* IEEE; 2022; 57–61.
Reference Source

30. Mitchell RE, Karchmer MA: **Chasing the Mythical Ten Percent: Parental Hearing Status of Deaf and Hard of Hearing Students in the United States.** *Sign Language Studies.* 2004; **4**(2): 138–163.
Publisher Full Text

31. Brown T, Mann B, Ryder N, *et al.*: **Language Models are Few-Shot Learners.** *arXiv preprint arXiv:2005.14165.* 2020.
Publisher Full Text

32. Asthana S, Hilleli S, He P, *et al.*: **Summaries, Highlights, and Action items: Design, implementation and evaluation of an LLM-powered meeting recap system.** *arXiv preprint arXiv:2307.15793.* 2023.
Publisher Full Text

33. Liu Y, Liu P, Radev DR, *et al.*: **BRIO: Bringing Order to Abstractive Summarization.** *arXiv preprint arXiv:2203.16804.* 2022.
Publisher Full Text

34. Roit P, Ferret J, Shani L, *et al.*: **Factually Consistent Summarization via Reinforcement Learning with Textual Entailment Feedback.** *arXiv preprint arXiv:2306.00186.* 2023.
Publisher Full Text

35. CORTEX2 Project: **CORTEX2 Architecture and Framework.** 2024.
Reference Source

36. Alcatel-Lucent Enterprise: **Rainbow Developer Portal.** 2024.
Reference Source

37. Lugaresi C, Tang J, Nash H, *et al.*: **MediaPipe: A Framework for Building Perception Pipelines.** *arXiv preprint arXiv:1906.08172.* 2019.
Publisher Full Text

38. Sanh V, Debut L, Chaumond J, *et al.*: **DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.** *arXiv preprint arXiv: 1910.01108.* 2019.
Publisher Full Text

39. Schmid P: *BART-Large CNN SamSum.* Hugging Face Transformers; 2021.
Reference Source

40. Tantaroudas ND, McCracken AJ, Karachalios I, *et al.*: INTERACT Pilot Study Data: Anonymised Questionnaire Responses and Performance Metrics. *Zenodo.* 2025; **2025**.
Publisher Full Text

41. Tantaroudas ND, McCracken AJ, Karachalios I, *et al.*: INTERACT ISL Gesture Dataset — International Sign Language Animation Data. *Zenodo.* 2025; **2025**.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ✓

Version 1

Reviewer Report 27 April 2026

https://doi.org/10.21956/openreseurope.25105.r71663

© **2026 Deepthi P.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✓ **PS Deepthi Deepthi**
Sharda University, Greater Noida, Uttar Pradesh, India

The authors present a framework by integrating Whisper-based speech recognition, NLLB multilingual translation, MediaPipe-driven ISL avatar animation, RoBERTa sentiment analysis, and BART meeting summarization for providing communication support for hard-of-hearing and deaf participants. The platform is an XR tool which integrates real-time speech-to-text transcription, International Sign Language (ISL) translation via 3D avatars, multilingual support, and sentiment analysis within immersive virtual environments. The paper is well written with sufficient description of the related works. The contributions are clearly mentioned in the introduction. The methodology is well explained with the support of pilot study and results. The ethical and consent statement is also included. The source code of the software and open access models used are also mentioned. Overall, the work presents a framework and its validation for supporting deaf people thereby enhancing their interactivity. The paper can be accepted in its current form.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Machine learning, deep learning, computer vision

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 20 April 2026

https://doi.org/10.21956/openreseurope.25105.r71845

© 2026 Tavares C. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Célia Tavares
CEOS.PP ISCAP Polytechnic of Porto, Porto, Portugal

The manuscript presents a very interesting and socially relevant contribution, as it has the potential to benefit deaf, hard-of-hearing, and multilingual users in remote communication settings.
The authors attempt to create a single communication ecosystem, and this modular architecture is one of the strongest aspects of the manuscript because it shows a possible path for future extensibility.
However, the manuscript has some methodological fragilities. The validation evidence is too limited to support the strength of some claims. The entire pilot study is based on ten participants across two demos, with six technical experts in the first and four deaf-community stakeholders in the second. This sample is extremely small. With such a small and mixed sample, descriptive statistics like “92% user satisfaction” and “4.6/5.0 overall experience” should be treated as preliminary observations, not as evidence of effectiveness. Furthermore, statements such as “Across both workshops, 40% of participants had no prior VR/XR experience, 40% had used VR once or twice, and 20% reported moderate experience,” while not formally incorrect, may convey a misleading sense of precision or generalizability given the small sample size (n = 10). A clearer and more appropriate approach would be to report absolute numbers, optionally accompanied by percentages in brackets, like the authors do in table 3, 4 and 5.
A closely related issue is that many performance claims are insufficiently operationalised. The manuscript reports “85%+ transcription accuracy,” “90% sentiment detection precision,” and “92% user satisfaction,” yet it does not define these measures with enough precision. For transcription accuracy, it is not clear whether the metric is word error rate, character error rate, sentence accuracy, or a more informal assessment. The discussion of sentiment accuracy also seems to rely partly on participants’ subjective impressions rather than on a formal evaluation. The same ambiguity appears in translation quality, where BLEU is mentioned in passing but no actual scores, test set, or comparative benchmark are reported. As a result, the reader cannot properly assess how strong the technical performance really is.
The sign-language component, which is the manuscript’s most novel feature, also shows some

fragilities. The manuscript states that the system uses a 747-video ISL corpus and a 750-sign vocabulary, with dictionary lookup plus fingerspelling fallback. This is a plausible prototype approach, but the manuscript does not provide a rigorous linguistic evaluation of the generated output. Instead, the evidence is mostly indirect, relying on questionnaire responses such as whether the avatar translation “felt accurate and natural.” Those subjective impressions are valuable, but they are not sufficient to support strong claims about sign-language translation quality.
Furthermore, the conclusions suggest that the system demonstrates readiness for expanded deployment and may transform accessibility in workplaces, schools, and cultural institutions. Given the pilot-scale evidence, this rhetoric is a bit strong. The work is promising, but it is more accurate to describe it as an early-stage prototype with encouraging pilot feedback than as a validated, deployment-ready platform.
The manuscript also has editorial inconsistencies. Some figures should appear when mentioned, to improve readability, but instead, for example, page six mentions figures from 1 to 5, and the respective images are only shown on pages 7 and 8. Others don’t have enough visual quality (Figure 6). Table 1 is also problematic, since the column header appears to say “Gender,” but the values shown in that column actually correspond to prior VR/XR experience categories such as “None,” “Once or twice,” and “Moderate.”
The manuscript claims no personally identifiable information was collected, yet it also mentions consent for anonymised photographs used in dissemination. This is not necessarily contradictory, but it would benefit from a clearer explanation of how image-based dissemination was handled while preserving anonymity.
Overall, this is a promising and potentially impactful manuscript, but the issues outlined above should be addressed.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Translation; neural machine translation; large language models for translation and other linguistic tasks; translation assessment and evaluation; technical communication.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.