*Human Dynamics, Product Evaluation and Quality*, Vol. 212, 2026, 9–17
https://doi.org/10.54941/ahfe1007808

# Communication Quality Assessment of an Augmented Reality System for Remote Support in Intensive Care Medicine

**Simon König[1], Lucas Seider[1], and Claus Backhaus[1,2]**

[1]Center for Ergonomics and Medical Engineering, FH Münster University of Applied Sciences, 48565 Steinfurt, Bürgerkamp 3, Germany

[2]Institute for Psychology and Ergonomics, Technical University of Berlin, Marchstraße 23, 10587 Berlin, Germany

## ABSTRACT

Augmented reality (AR) systems can provide remote expert support for safety-critical intensive care workflows. Patient safety depends on reliable audiovisual communication; poor transmission may increase mental workload, reduce comprehension, and promote errors. Because systematic evidence on AR communication performance is limited, we evaluated an AR remote-support system and its clinical suitability. We assessed audio quality, end-to-end latency, and audiovisual synchrony using subjective and objective measures. In a user study (n = 10), participants listened to standardized speech under typical intensive-care background noise and rated intelligibility, naturalness, noise perception, listening effort, and overall quality on a five-point Mean Opinion Score (MOS). Latencies were measured with a test signal containing synchronous visual and acoustic markers; sender and receiver views were recorded and aligned using an atomic clock. Marker analysis yielded audio and video latencies under two network conditions; audiovisual synchrony was computed as their difference. Bitrate, packet loss, and jitter were logged and combined with latency measures into an objective MOS. Participants rated audio quality as good (MOS 4.28/5). Audio latencies were 149 (±91) ms and 167 (±60) ms; video latencies were 156 (±64) ms and 186 (±44) ms. Audiovisual synchrony was −10 (±76) ms and 38 (±65) ms. All latencies were below the 250 ms teleconsultation threshold, and synchrony remained within the −188 to 75 ms tolerance for action-oriented tasks, indicating no perceptible asynchronies. The objective MOS was 3.86, corroborating the subjective ratings. Overall, the system meets communication requirements for telemedical remote support and is suitable for pilot clinical deployment.

**Keywords:** Augmented reality (AR), Medical remote support, Communication quality assessment, Performance evaluation

## INTRODUCTION

Intensive care units (ICUs) are designed for the treatment and monitoring of life-threatening conditions. Advances in medical technology enable the care of increasingly severe and complex disease patterns. At the same time, growing technical complexity increases the susceptibility of clinical workflows to disruptions. This raises the need for rapidly available expertise for fault

Received February 4, 2026; Revised April 3, 2026; Accepted April 17, 2026; Available online July 20, 2026

diagnosis and troubleshooting. In clinical practice, such expertise is often not immediately available on site, making external consultations necessary.

Telemedical systems can reduce this availability gap by providing location-independent expert support. Augmented reality (AR) technologies are particularly suitable for this purpose. Compared with conventional video conferencing or telephone support, AR allows computer-generated content, such as visual markings or step-by-step instructions, to be overlaid directly in the user's field of view. This can support ICU staff during safety-critical workflows.

Previous studies demonstrate feasibility and acceptance of AR-based remote support. Palumbo (2022) reports broad applicability for teleconsultations. Dinh et al. (2023) consistently describe positive ratings regarding effectiveness, feasibility, and acceptance. For training-related scenarios, Wolf et al. (2021) showed a reduction in knowledge-based errors using AR-based instructions. However, safe clinical use depends not only on basic functionality and usability. It also depends on the quality of audiovisual communication.

Insufficient audio quality reduces the intelligibility of critical information and can promote information loss and incorrect actions (Patel et al., 2022). Independently, impaired sensory integration, for example due to high latency or insufficient audio-video synchrony, may increase cognitive load and prolong reaction times (Väljamäe, 2005).

Therefore, this study evaluates the communication quality of an AR system for telemedical remote support and assesses its suitability for intensive care. Audio quality, latency, and audiovisual synchrony are captured using subjective and objective measures, combined into an overall assessment, and interpreted using established metrics.

## MATERIALS

A self-developed AR system for remote support in intensive care was investigated. The system consists of an AR head-mounted display (HoloLens 2, Microsoft Corp., Redmond, WA, USA) used by ICU staff and a tablet computer (Galaxy Tab S8, Samsung Electronics Co., Ltd., Suwon, Republic of Korea) used by the remote expert.

The implementation is based on a Web Real-Time Communication (WebRTC) application and enables bidirectional audiovisual communication. In addition, the remote expert can place visual annotations via the tablet into the HMD wearer's field of view to provide context-specific guidance. Figure 1 illustrates the system principle.

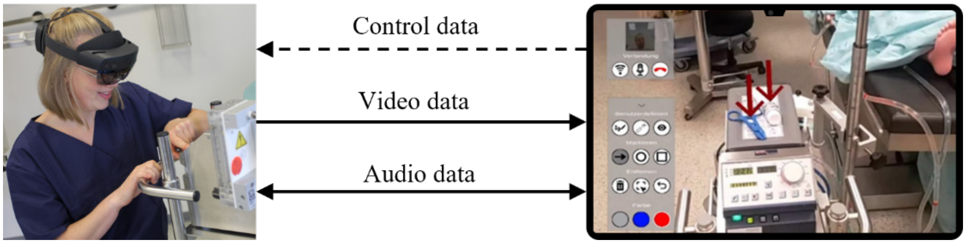

**Figure 1:** Schematic data flow of the AR system between ICU staff (HoloLens 2) and remote expert (tablet): Video is streamed from HoloLens 2 to tablet, audio is exchanged bidirectionally, and control data (e.g. annotations) are sent from the tablet to the HoloLens 2.

## METHODS

### Subjective Assessment of the Audio Quality

Subjective audio quality was assessed in a user study with 10 participants. Participants used the tablet to conduct a standardized conversation with the experimenter, who wore the HoloLens 2. Both were located in separate rooms and could communicate only via the AR system. In the experimenter's room, an ICU-typical acoustic background was simulated.

After the conversation, participants rated audio quality using the Mean Opinion Score (MOS; ITU, 2023). The MOS covered overall speech transmission quality, intelligibility, voice naturalness, perceived noise interference, and listening effort. Ratings were given on a five-point Likert scale (Table 1).

**Table 1:** Five-point Mean Opinion Score (MOS) scale used for the subjective audio quality assessment.

| MOS value | Rating | Description |
|---|---|---|
| 5 | Excellent | No noticeable impairments |
| 4 | Good | Minor impairments |
| 3 | Moderate | Noticeable disturbances |
| 2 | Poor | Significant impairments |
| 1 | Useless | Communication not possible |

Mean values were computed for each dimension and as an overall mean across dimensions.

### Objective Assessment of Latencies and Audiovisual Synchrony

Audiovisual synchrony was defined as the temporal offset between video and audio. It was computed as the difference between video transmission time ($T_V$) and audio transmission time ($T_A$), expressed as $\Delta t = T_V - T_A$. A positive value of $\Delta t$ indicates that the video signal was delayed relative to the audio signal (audio lead), while a negative value indicates that the audio signal was delayed relative to the video signal (video lead).

Measurements were conducted in two separate rooms with the HoloLens 2 as sender and the tablet as receiver. A standardized test video with synchronized visual (white-light flash) and acoustic markers (3-kHz tone) served as the reference signal. Sender and receiver views were recorded using a camera (GoPro Hero 11, GoPro Inc., San Mateo, CA, USA). A web-based atomic clock was recorded in both rooms to align the recordings temporally. Figure 2 shows the setup.

Two network configurations were tested: (1) WLAN–WLAN, with both devices connected to the same Wi-Fi network as a reference for favorable conditions; and (2) WLAN–4G, with the HoloLens on Wi-Fi and the tablet on a cellular network, representing an application-oriented scenario.

**Figure 2**: Experimental setup for measuring audio and video latency and audiovisual synchrony using synchronized light and tone makers and time alignment via web-based atomic clock.

The protocol covered five test days with three connection runs per day (morning, noon, evening). For each run, 50 synchronization impulses were recorded per network configuration. Analysis was automated in Python. Tone markers were detected via frequency analysis; light impulses were identified via luminance changes. Audio latency, video latency, and Δt were then computed and interpreted using established tolerance ranges for telemedical applications (Eg & Behne, 2015; Nankaku et al., 2022; Korai et al., 2023). Table 2 summarizes the acceptance ranges.

**Table 2**: Acceptance ranges for audio and video latency and audiovisual synchrony by application class (latency: telesurgery and teleconsultation; synchrony: action-oriented tasks, speech-oriented tasks).

| Application | Latency | Synchrony |
|---|---|---|
| Telesurgery | ≤ 100 ms | |
| Teleconsultation | ≤ 250 ms | |
| Action-oriented tasks | | $-188 \leq \Delta t \leq 75$ ms |
| Speech-oriented tasks | | $-258 \leq \Delta t \leq 131$ ms |

### Objective Overall Assessment

A parametric MOS model for interactive audiovisual systems was used to derive an objective MOS (ITU, 2025). In parallel to the measurements above, network and media parameters were logged, including jitter and packet loss rates for audio and video. In addition, audio and video bitrate, video frame rate, and resolution parameters were documented. Measured audio and video latencies ($T_A$, $T_V$) were also included.

The model first computes separate quality scores for audio ($Q_A$) and video ($Q_V$). Video quality is based on a device-specific baseline under ideal conditions and is reduced according to network impairments. Audio quality is modeled via a normalized quality factor that accounts for codec-dependent properties and packet loss, using an equipment impairment factor ($I_E$) and a packet-loss robustness factor ($B_{pl}$). For the Opus codec, parameter values were taken from the literature (Al-Ahmadi et al., 2020; Michael et al., 2021).

$Q_A$ and $Q_V$ are then fused into an audiovisual quality score ($Q_{AV}$), reflecting cross-modal interaction in perception. Delay effects are subsequently modeled: interaction delay ($f_{delay}$) and audiovisual synchrony ($f_{sync}$) enter as penalty terms and reduce overall quality ($Q_{VT}$). Model coefficients were taken from reference tables (ITU, 2025). The resulting objective MOS was interpreted on the same five-point MOS scale as the subjective ratings (Table 1).

## RESULTS

### Subjective Assessment of the Audio Quality

Participants rated the AR system with an overall MOS of 4.28, corresponding to good audio quality. Mean scores for individual dimensions ranged from 3.8 to 4.8, corresponding to good to excellent ratings. Table 3 summarizes the results.

**Table 3:** Subjective audio quality rating (MOS) by dimension and overall score.

| Quality Dimension | Mean Value | Rating |
|---|---|---|
| Voice transmission | 4.1 | Good |
| Intelligibility | 4.6 | Good |
| Naturalness | 3.8 | Good |
| Noise interferences | 4.1 | Good |
| Effort | 4.8 | Excellent |
| Overall | 4.28 | Good |

### Objective Assessment of Latencies and Audiovisual Synchrony

Measured latencies for both network configurations were within the acceptable range for telemedical consultations (≤ 250 ms). Requirements for telesurgical applications (≤ 100 ms) were not met. Audiovisual synchrony remained within acceptance ranges for both speech- and action-oriented tasks. Table 4 shows the results.

**Table 4:** Audio latency, video latency, and audiovisual synchrony under WLAN-WLAN and WLAN-4G conditions (mean ± SD).

| Parameter | Symbol | WLAN-WLAN | WLAN-4G |
|---|---|---|---|
| Video latency | TV | 156.3 ± 63.6 ms | 186.6 ± 44.1 ms |
| Audio latency | TA | 166.6 ± 59.9 ms | 149.0 ± 90.7 ms |
| Synchrony | Δt | – 10.3 ± 76.0 ms | 37.6 ± 64.6 ms |

## Objective Overall Assessment

The objective overall assessment was based on continuously logged network/media parameters and the measured audio/video latencies. Input variables are summarized in Table 5.

**Table 5:** Logged network and media parameters used as inputs for the objective MOS model, including measured latencies and derived penalty terms.

| Parameter | Symbol | Value |
|---|---|---|
| Video bit rate | $Br_V$ | 2,415.56 kbit/s |
| Video frame rate | $Fr_V$ | 30 fps |
| Video packet loss rate | $Plr_V$ | 0.17 % |
| Audio packet loss rate | $Plr_A$ | 2.56 % |
| Jitter | jit | 15 ms |
| Audio latency | $T_A$ | 149.0 ms |
| Video latency | $T_V$ | 186.6 ms |
| Interaction delay penalty term | $f_{delay}$ | 0.0059 |
| Audiovisual synchrony penalty term | $f_{sync}$ | 0.0007 |

Based on these inputs, $Q_A$ and $Q_V$ were calculated, fused into $Q_{AV}$, and then reduced to $Q_{VT}$ using penalty terms for interaction delay and audiovisual synchrony. The resulting MOS-based values are presented in Table 6.

**Table 6:** Objective quality scores ($Q_A$, $Q_V$, $Q_{AV}$, $Q_{VT}$) and corresponding MOS-based interpretation.

| Quality dimension | Symbol | Value | Rating |
|---|---|---|---|
| Audio quality | QA | 4.28 | Good |
| Video quality | QV | 3.36 | Acceptable |
| Audiovisual quality | QAV | 3.51 | Good |
| Overall quality | QVT | 3.86 | Good |

Video quality was in the acceptable range. Audio quality, audiovisual quality, and overall quality were in the good range. Overall, the objective evaluation indicated good to acceptable communication quality for the AR system.

## DISCUSSION

AR-based remote support represents a relevant use case in intensive care, particularly for time-critical interventions (Dinh et al., 2023). The results indicate that the investigated AR system provides communication quality that is suitable for this purpose.

High subjective audio quality is clinically relevant. ICUs are characterized by high alarm densities and elevated sound pressure levels (Backhaus et al., 2023). Both factors can impair information processing (Poncette et al., 2021; Sangari et al., 2021). Background noise further increases listening and processing effort (Wendt et al., 2016). Against this background, low perceived noise interference and low listening effort can be regarded as prerequisites for sustained, low-burden communication.

Mean audio and video latencies were within ranges reported as tolerable for teleconsultation. Mean audiovisual synchrony was also within published tolerance windows (Eg & Behne, 2015; Nankaku et al., 2022; Korai et al., 2023). For safety-critical communication, variance is additionally important. Individual latency peaks can disrupt interaction even when mean values remain unremarkable.

The objective MOS was lower than the subjective audio MOS. This is expected because the objective overall score integrates multiple components, including penalty terms for delay and audiovisual synchrony. This aligns with evidence that audio and visual perception are not independent and that task demands can shift the weighting of modalities (Lee et al., 2009; Becerra Martínez et al., 2021).

Limitations primarily concern generalizability. Experiments were conducted in a laboratory environment. Real ICU settings show dynamic soundscapes and variable network utilization. In addition, the sample size was small and below recommended sizes for MOS studies (ITU, 2014).

Future studies should evaluate communication quality under real ICU conditions. Initial clinical use should be limited to non-critical events to avoid compromising patient safety and process safety.

## CONCLUSION

Overall, the combination of good subjective audio quality, low latencies, uncritical audiovisual synchrony, and a good objective MOS supports the system's suitability as a technical basis for telemedical remote support. A time-limited pilot deployment on an ICU is a reasonable next step. Key criteria should include stability of communication quality under real network load, compliance with teleconsultation latency thresholds during load peaks, and avoidance of critical outliers in action-oriented situations.

## REFERENCES

Al-Ahmadi, M., Pocta, P. and Melvin, H. (2020). “Derivation of E-model Equipment Impairment Factors for Narrowband and Wideband Opus Codec Using the Instrumental Method”, proceedings of the 31st Irish Signals and Systems Conference (ISSC). IEEE. pp. 1–6. DOI: 10.1109/ISSC49989.2020.9180160.

Backhaus, C., Carl, A., Siebers, S. and Urlichs, F. (2023). “Lärmbelastung einer neonatologischen Intensivstation: Gibt es einen Tag-Nacht-Unterschied der Schallexposition?”, Zentralblatt für Arbeitsmedizin, Arbeitsschutz und Ergonomie, Volume 73. pp. 24–31. DOI: 10.1007/s40664-022-00486-6.

Becerra Martínez, H., Hines, A. and Farias, M.C.Q. (2021). “Perceptual Quality of Audio-Visual Content with Common Video and Audio Degradations”, Applied Sciences, Volume 11. 5813. DOI: 10.3390/app11135813.

Dinh, A., Yin, A.L., Estrin, D., Mishra, N., Weng, Y., Zhan, C., Nikzad, N., Ramirez, E., Sarrafzadeh, M. and Naeim, A. (2023). “Augmented reality in real-time telemedicine and telementoring: Scoping review”, JMIR mHealth and uHealth, Volume 11. 5813. DOI: 10.2196/45464.

Eg, R. and Behne, D.M. (2015). “Perceived synchrony for realistic and dynamic audiovisual events”, Frontiers in Psychology, Volume 6. 736. DOI: 10.3389/fpsyg.2015.00736.

International Telecommunication Union – Telecommunication Standardization Sector (ITU-T). (2014). Recommendation ITU-T P.913: Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment. Geneva: ITU.

International Telecommunication Union – Telecommunication Standardization Sector (ITU-T). (2023). Recommendation ITU-T P.910: Subjective video quality assessment methods for multimedia applications. Geneva: ITU.

International Telecommunication Union – Telecommunication Standardization Sector (ITU-T). (2025). Recommendation ITU-T P.940: Computational model used for the monitoring and quality assessment of videotelephony services. Geneva: ITU.

Korai, T., et al. (2023). “Development of an ultralow-latency communication system using shared internet access: Promoting the implementation of telemedicine”, Surgery Today, Volume 53 No. 3. pp. 522–525. DOI: 10.1007/s00595-022-02602-8.

Lee, J.-S., De Simone, F. and Ebrahimi, T. (2009). “Influence of audio-visual attention on perceived quality of standard definition multimedia content”, proceedings of the 2009 International Workshop on Quality of Multimedia Experience (QoMEX). DOI: 10.1109/QOMEX.2009.5246983.

Michael, T., Mittag, G. and Möller, S. (2021). “Derivation of Equipment Impairment Factors from ITU-T P.863 for Fullband Speech Codecs”, proceedings of DAGA 2021 – 47. Jahrestagung für Akustik, Wien. pp. 1273–1276.

Nankaku, A., et al. (2022). “Maximum acceptable communication delay for the realization of telesurgery”, PLOS ONE, Volume 17 No. 10. e0274328. DOI: 10.1371/journal.pone.0274328.

Palumbo, A. (2022). “Microsoft HoloLens 2 in Medical and Healthcare Context: State of the Art and Future Prospects”, Sensors, Volume 22 No. 20. 7709. DOI: 10.3390/s22207709.

Patel, V., Li, C.H., Rye, V., Entezami, P., Mondal, T., Liu, S., Ginsburg, S.B., Cicchetti, M.G., Andersen, D.K., Bartlett, D.L., Zureikat, A.H., Zeh, H.J., Tsung, A., Lee, K.K.W. and Hogg, M.E. (2022). “A comparison of WebRTC and conventional videoconferencing for synchronized remote medical image presentation”, Journal of Imaging Informatics in Medicine, Volume 35. pp. 665–672. DOI: 10.1007/s10278-021-00544-0.

Poncette, A.-S., Wunderlich, M.M., Spies, C., Heeren, P., Vorderwülbecke, G., Salgado, E., Kastrup, M., Feufel, M.A. and Balzer, F. (2021). "Patient Monitoring Alarms in an Intensive Care Unit: Observational Study With Do It-Yourself Instructions", Journal of Medical Internet Research, Volume 23 No. 5. e26494. DOI: 10.2196/26494.

Sangari, A., Emhardt, E.A., Salas, B., Avery, A., Freundlich, R.E., Fabbri, D., Shotwell, M.S. and Schlesinger, J.J. (2021). "Delirium Variability is Influenced by the Sound Environment (DEVISE Study): How Changes in the Intensive Care Unit Soundscape Affect Delirium Incidence", Journal of Medical Systems, Volume 45 No. 8. 76. DOI: 10.1007/s10916-021-01752-5.

Väljamäe, A., Kohlrausch, A., van de Par, S., Larsson, P., Västfjäll, D. and Kleiner, M. (2005). "Audio-visual interaction and synergy effects: implications for cross-modal optimization of virtual and mixed reality applications", Handbook of Presence (Draft version 050523). Available online: https://www.researchgate.net/publication/259463626 (last access: 29.01.2026).

Wendt, D., Dau, T. and Hjortkjær, J. (2016). "Impact of Background Noise and Sentence Complexity on Processing Demands during Sentence Comprehension", Frontiers in Psychology, Volume 7. 345. DOI: 10.3389/fpsyg.2016.00345.

Wolf, J., Wolfer, V., Halbe, M., Heinemann, A., Müller, M., Schmitz, B., Reuter, D.A., Mommsen, P. and Krettek, C. (2021). "Comparing the effectiveness of augmented-reality-based and conventional instructions during single ECMO cannulation training", International Journal of Computer Assisted Radiology and Surgery, Volume 16. pp. 1171–1180. DOI: 10.1007/s11548-021-02408-y.