Interoperable Total Conversation

Matjaž Debevc
Faculty of Electrical Engineering and Computer Science
University of Maribor
Maribor, Slovenia
matjaz.debevc@guest.um.si

Gunnar Hellström
GHAccess
Haninge, Sweden
gunnar.hellstrom@ghaccess.se

Olumide
Ex2 Management Consulting Ltd
United Kingdom
jogunbekun@yahoo.com

Mike Pluke
Castle Consulting
United Kingdom
mike.pluke@castle-consult.com

Agata Sawicka
ITS
Sweden
agata@universal-ict.eu

Abstract

The European Accessibility Act (EAA) requires that wherever real-time video communication is provided, Total Conversation must be supported. Total Conversation enables simultaneous exchange of real-time text (RTT), video and voice, providing inclusive and flexible communication for all. This paper discusses the work of the project ETSI STF 674 and its development of the ETSI ES 204 009 standard on Interoperable Total Conversation. It presents performance requirements for each media modality, integration with assistive technologies, test methods for video quality, and the importance of interoperability and relay services. The implementation of Total Conversation ensures compliance with EU accessibility laws and enhances usability in a broad range of communication contexts.

Keywords

Total Conversation, Real-Time Text, Videoconference, Voice Communication, Interoperability, Accessibility, Relay Services

1 Introduction

Inclusive and accessible communication technologies are not only a matter of equity but are increasingly becoming a legal obligation across the European Union. The European Accessibility Act (EAA), adopted in 2019, mandates that certain digital products and services - including those used for communication - must be accessible to all individuals, regardless of their disabilities. A central requirement of the EAA is the support for Total Conversation (TC) in any context where real-time video communication is offered. This includes customer service platforms, emergency communications, conferencing systems, and any ICT-based interpersonal communication service.

Information Society 2025, 6–10 October 2025, Ljubljana, Slovenia

http://doi.org/DOI_RECEIVED_AFTER_REVIEW

Total Conversation is defined as the simultaneous, synchronized provision of three media modalities - real-time text (RTT), video and voice. When fully integrated, these media support diverse communication preferences and needs, such as sign language over video, speech supported by lip-reading, and typed text for individuals who are deaf, hard of hearing, deafblind, or have speech and cognitive impairments. Rather than treating these communication modes as separate alternatives, TC ensures they are available concurrently and interoperable, enabling flexible and responsive conversations adapted to the users' abilities.

While individual components like video calls or text chat systems are widely deployed, the combination of modalities in a synchronized and interoperable manner has not been consistently implemented across platforms and services. The lack of interoperability also hinders effective communication between different systems, vendors, and devices - an issue particularly problematic in emergency contexts, transnational communication, and public services.

To address this, ETSI TC Human Factors (HF) initiated multiple projects aimed at translating the legal obligations of the EAA into technical requirements. One of these, Special Task Force 674 (STF 674), is responsible for developing the standard ETSI ES 204 009 [1], which defines the functional, media, and interoperability requirements for Total Conversation systems. It complements other standards work such as ETSI EN 301 549 (accessibility requirements for ICT products and services) [2] and ETSI ES 202 975 (relay service specifications) [5], forming a coherent standards ecosystem aligned with regulatory and user needs.

This paper provides an overview of the ETSI ES 204 009 standard, elaborates on the minimum performance levels necessary for usable TC services, presents easy-to-implement test methodologies to verify conformance, and highlights the importance of ensuring interoperability across services. It also explores integration with assistive technologies such as hearing aids and relay services and outlines practical scenarios where TC brings significant value.

2 Components of Total Conversation

Total Conversation is built upon three core media streams (real-time text, video and voice) that are delivered in parallel and are functionally integrated between two or more user equipment (Fig. 1):

- **Real-Time Text (RTT):** Unlike instant messaging or email, RTT allows users to transmit each character as it is typed, creating a live and dynamic text flow. This enables tight interaction, clarifications, and even interruption—key aspects of natural conversation.
- **Video:** This channel primarily serves sign language communication, non-verbal cues such as facial expressions and gestures, and lip-reading for users with residual hearing. It plays a critical role for the deaf and hard-of-hearing community and must be of sufficient quality to support complex sign articulation and facial grammar.
- **Voice:** The voice channel remains vital for many users, including those with visual or cognitive disabilities. Wideband audio enhances intelligibility and makes communication more natural, especially in challenging environments or over mobile networks.

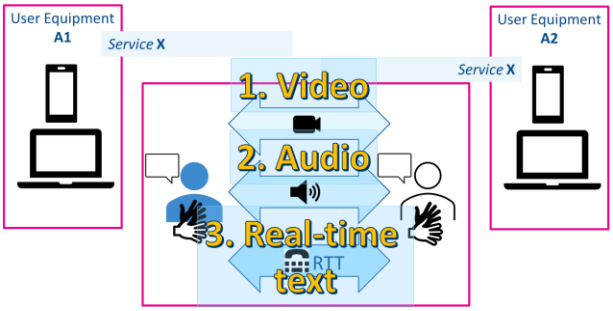

Figure 1: Components of Total Conversation

The value of TC lies not only in providing these media independently but in combining them in a way that allows users to shift seamlessly between modalities or use them concurrently. This flexibility is essential in mixed-ability communication scenarios and ensures redundancy if one channel is impaired [6].

3 Interoperability considerations

The effectiveness of Total Conversation hinges not only on the quality of individual media streams but also on their interoperability across platforms, networks, and devices. Interoperability is essential to ensure that communication can occur seamlessly between users regardless of their service provider, geographic location, or technical infrastructure. Without interoperability, the fragmentation of services would result in digital exclusion, especially in scenarios requiring urgent or cross-border communication.

The ETSI ES 204 009 standard defines a set of interoperability requirements that enable real-time text, video, and voice to be exchanged using established communication protocols. It promotes the use of SIP (Session Initiation Protocol), IMS (IP Multimedia Subsystem), and WebRTC (Web Real-Time Communication) as foundational signaling and media transport frameworks. These technologies support dynamic media negotiation, allowing devices and services to determine the most appropriate encoding and channel configurations during call setup. This ensures that participants can successfully exchange synchronized real-time text, video, and voice, regardless of the underlying system implementations.

Media stream compatibility is ensured through alignment on common codecs and formats. For voice, wideband codecs such as AMR-WB and Opus are recommended due to their robustness in variable bandwidth conditions and their superior audio clarity. For video, H.264 is a widely supported codec that balances quality and compression efficiency, while for real-time text, the standardized RTP payload format defined in RFC 4103 is used [4], allowing immediate character-by-character transmission over IP networks.

Interoperability also depends on effective session negotiation. Using Session Description Protocol (SDP), endpoints can exchange capabilities such as supported codecs, resolution, frame rates, and synchronization tolerances. The ETSI ES 204 009 ensures that these mechanisms are harmonized, reducing the risk of call failures or degraded media performance due to incompatible configurations.

A significant challenge in TC is achieving interoperability not only between devices of the same type but across heterogeneous systems. For example, a call initiated from a WebRTC browser must be able to connect with a SIP-based desktop phone or a mobile application, with full support for all three TC modalities. This requires robust gateway implementations and adherence to interface profiles defined in the standard.

Furthermore, the standard addresses interoperability in addressing and routing. It supports both E.164 telephone numbers and URI-based identifiers (such as SIP URIs), making it possible to contact TC endpoints from traditional voice networks as well as IP-based applications. This flexibility is particularly crucial for integrating TC into public services such as emergency response systems and governmental communication platforms.

Cross-vendor interoperability testing is a vital implementation step. The ETSI ES 204 009 encourages conformance testing, reference implementations, and public testing events to validate interoperability among different vendors and service providers.

4 Performance requirements

To be effective, Total Conversation must adhere to strict quality of service (QoS) criteria that support human factors in communication. Video must provide a minimum spatial resolution of QVGA (320x240 pixels), a frame rate of at least 20 fps, and end-to-end latency below 400 msec. These specifications ensure that fast gestures and subtle facial expressions remain discernible and are not lost due to compression or jitter. In practical terms, this means that users engaging in sign language conversations must be able to see hand shapes, movements, facial expressions, and gaze direction with clarity. Loss of visual fidelity can significantly impair the grammatical and semantic meaning of sign language.

Real-Time Text must be delivered with under 1 second of delay, support at least 30 characters per second throughput, and be fully compliant with Unicode for multilingual and symbol-rich communication. This allows for high responsiveness, including overlapping dialogue, corrections, and incremental

feedback, which are vital in dynamic conversation. Inadequate RTT performance can lead to conversational breakdowns and exclusion from interactive discussions.

Voice should operate within a frequency band of 250 Hz to 7,000 Hz with minimal compression artifacts, ensuring the intelligibility of a wide range of vocal tones, especially for users with hearing impairments who rely on residual hearing or hearing aids. Latency should be maintained below 400 milliseconds to avoid echoing, speaker overlapping, and degraded user experience, particularly in multilingual or moderating calls.

Synchronization of video and audio must stay within ±100 msec to preserve lip-sync, which is critical for users who lip-read or use speech-reading techniques. Asynchrony between modalities reduces the effectiveness of multimodal communication and may result in reduced comprehension or fatigue over long calls.

These values were derived from user studies, practical experience, and ITU-T recommendations, particularly H-Series Supplement 1, which focuses on the technical requirements that support the perceptual needs of users engaging in sign language and lip-reading communication [3]. Compliance with these parameters ensures not only legal adherence but also a user experience that supports autonomy, agency, and inclusion.

5 Integration with assistive technologies

For Total Conversation to be truly inclusive, it must work in harmony with the wide variety of assistive technologies used by people with disabilities. The effectiveness of TC is amplified when the real-time text, video, and voice channels are compatible with user-specific devices and support tools that enhance communication based on individual needs.

One of the most critical integrations is with **hearing aids and cochlear implants**, which often rely on Bluetooth connectivity or telecoil (T-coil) systems to receive audio directly from digital devices. Emerging technologies such as Auracast™ Bluetooth LE Audio allow group broadcasting, making it possible to seamlessly connect to public or shared audio systems, such as in classrooms or conference settings. Ensuring low-latency, high-fidelity audio from the TC system into these devices significantly improves the accessibility of spoken communication.

For users who are blind or deafblind, **screen readers and Braille displays** provide essential access to real-time text. TC systems must ensure that RTT content is presented in a way that is compatible with screen reading software, with proper semantic structure, keyboard navigation, and timely character transmission.

Users with complex communication needs may depend on **Augmentative and Alternative Communication (AAC) devices**, which enable them to construct and transmit messages through symbol boards, text-to-speech systems, or visual interfaces. A well-designed TC service allows for these devices to input into the voice channel via synthesized speech, or directly into the RTT stream, without introducing latency or requiring manual transcription.

Moreover, TC systems must be controlled via **accessible user interfaces** that meet the requirements of the ETSI EN 301 549. This includes support for keyboard navigation, voice commands, magnification, high-contrast modes, and alternative input devices. System setup, call controls, and media selection must be operable without vision, hearing, or precise motor control.

Beyond hardware and software compatibility, **personalization** plays a key role. Users should be able to configure the prominence and layout of modalities based on their preferences—such as enlarging the video feed for sign language, prioritizing RTT over voice, or adjusting caption position and font.

6 Relay services and modality conversion

Relay services are essential components of the Total Conversation ecosystem, enabling accessible communication between users who rely on different modalities. These services function as real-time intermediaries that convert between speech, text, and sign language, ensuring that all participants can engage effectively, regardless of their abilities or communication preferences.

Video Relay Services (VRS) connect sign language users with hearing individuals by incorporating a live sign language interpreter into the communication loop. The interpreter receives the video stream from the signing user, translates it into spoken language, and transmits it to the hearing person, while simultaneously converting spoken replies into sign language. VRS is particularly important for users who communicate primarily through national or regional sign languages, and it requires high-quality video transmission to support accurate interpretation.

Text Relay Services (TRS) enable users who communicate via text to interact with voice telephone users. This can be facilitated by a human relay operator who voices the typed messages and types back the spoken responses, or by automated systems with speech-to-text and text-to-speech capabilities. TRS must support real-time interaction, minimize delays, and protect user privacy and data.

Captioned Telephony is another form of relay service that provides a live transcription of the spoken content during a voice call. It benefits users who can speak but have difficulty hearing, by overlaying real-time text captions of the other party's speech. These captions are typically generated by either human captioners or automated speech recognition systems and are displayed synchronously with the voice channel.

The effectiveness of relay services depends on several factors, including low latency, high accuracy, robust data security, and compliance with accessibility and privacy regulations. The ETSI ES 202 975 standard outlines the functional and quality requirements for implementing reliable and interoperable relay services within the context of Total Conversation.

Furthermore, modality conversion must be seamless, meaning that users should not experience noticeable interruptions or mismatches when transitioning between modes. The system should support smooth transfers between modalities within the same call session, for example, shifting from text to video or adding a voice channel, without requiring disconnection or renegotiation.

7 Application scenarios

Total Conversation has widespread applicability and supports full participation in many areas of modern life by making real-time communication accessible and inclusive.

In **emergency communications**, Total Conversation enables direct, multimodal contact with public safety answering points (PSAPs). A deaf or speech-impaired individual can initiate a call using real-time text, video, and voice as needed. Sign language users can communicate with interpreters, while text can provide clarity or serve as a fallback. Visual information allows emergency personnel to assess scenes, verify distress signals, and provide visual guidance - capabilities impossible with voice-only communication.

In **education and e-learning**, TC creates accessible virtual classrooms and lecture environments. Deaf and hard-of-hearing students can follow spoken content via real-time captions or sign language and interact via RTT. Teachers can use TC platforms to provide differentiated instruction tailored to communication preferences. Moreover, students with cognitive or speech disabilities can participate using AAC devices integrated into TC sessions.

Telehealth and remote care benefit from TC through improved doctor-patient interactions, especially where patients have sensory or speech impairments. Patients can describe symptoms using video, confirm prescriptions in text, and receive visual reassurance. For mental health services, TC supports rapport-building with facial expressions, tone of voice, and real-time emotional feedback—all essential elements in therapeutic contexts.

In **professional collaboration**, TC bridges communication gaps in mixed-ability work environments. It ensures that meetings, interviews, and team discussions include all participants equally. For example, a deaf employee can engage via sign language while receiving spoken input through captions or RTT. The combination of modalities reduces misunderstandings and supports accurate documentation of decisions.

Social inclusion is another vital domain. Total Conversation allows family members with different abilities to stay connected in a rich and natural way. Elderly individuals with hearing loss, young children learning to type, and multilingual family members can all participate in one conversation using the mode that suits them best. Social services and community events can also be made accessible through TC, fostering engagement and civic participation.

As the digital landscape evolves toward hybrid, remote, and inclusive communication, Total Conversation stands out as a technology that supports equity, autonomy, and human connection. By enabling flexible and barrier-free participation in daily life, it reinforces the right to communicate for all.

8 Conclusion

Interoperable Total Conversation is a cornerstone for inclusive communication in the digital age. By integrating synchronized real-time text, video, and voice and aligning with the European Accessibility Act, it enables a communication paradigm where everyone - regardless of sensory, cognitive, or linguistic ability - can fully participate in society. The ETSI ES 204 009 standard, developed by STF 674, provides a practical foundation for achieving this interoperability and accessibility across diverse systems and platforms.

The implementation of TC contributes to a communication infrastructure. It ensures that users can engage in both casual and critical conversations - from contacting emergency services to participating in professional meetings or accessing telehealth consultations - using the modality that best suits their needs at any moment.

Furthermore, Total Conversation is more than a technological solution; it reflects a broader societal shift toward inclusivity and digital equity. By embracing standardized performance metrics, interoperability protocols, and integration with assistive technologies, TC systems can reduce barriers, enhance quality of life, and support equal opportunities in education, employment, healthcare, and beyond.

Going forward, continued collaboration among policymakers, standardization bodies, service providers, and accessibility advocates will be essential to scale adoption, ensure consistent implementation, and refine technical solutions. As digital communication evolves, Total Conversation should remain a guiding principle for designing accessible, future-proof communication environments that serve everyone. With ETSI ES 204 009 and aligned standards, service providers can implement reliable, high-quality, and accessible real-time communication. The success of TC depends on adherence to technical standards, support for assistive technologies, and a commitment to cross-platform interoperability.

Acknowledgments

The authors gratefully acknowledge the support and contributions of the European Telecommunications Standards Institute (ETSI), in particular the Specialist Task Force STF 674 (Ref. Body TC HF), whose work on the ETSI ES 204 009 standard on Interoperable Total Conversation served as the foundation for this paper. Authors acknowledge that this article was partially generated by ChatGPT (powered by OpenAI's language model, GPT-4.5; http://openai.com). The editing was performed by human authors.

References

[1] ETSI ES 204 009 V1.1.1 (2025-06). Human Factors (HF); Requirements for interoperable total conversation services.
[2] ETSI EN 301 549 V3.2.1 (2021-03). Accessibility requirements for ICT products and services
[3] ITU-T H-Series Supplement 1 (1999). Application profile - Sign language and lip-reading real-time conversation
[4] IETF RFC 4103. RTP Payload for Text Conversation
[5] ETSI ES 202 975. Requirements for relay services
[6] IETF RFC 9071. RTP-Mixer Formatting of Multiparty Real-Time Text