Final draft ETSI ES 204 009 V1.1.1 (2025-06)

Human Factors (HF);
Requirements for interoperable total conversation services

Reference
DES/HF-00301559

Keywords
accessibility, HF, ICT, procurement, relay, service, total conversation

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI IPR online database.

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™**, **LTE™** and **5G™** logo are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM®** and the GSM logo are trademarks registered and owned by the GSM Association.

BLUETOOTH® is a trademark registered and owned by Bluetooth SIG, Inc.

Foreword

This final draft ETSI Standard (ES) has been produced by ETSI Technical Committee Human Factors (HF), and is now submitted for the ETSI Membership Approval Procedure (MAP).

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

1 Scope

The present document provides detailed functional, service and accessibility requirements from the Human Factors perspective for the implementation of and compliance with the European Accessibility Act [i.27] which requires the provision of interoperable total conversation services wherever voice and video communication is available. It details how to provide an interoperable total conversation service, with voice, real-time text and sign language-supportive video communication media synchronized, as a context-optimized service, with no need for additional pre-registration. It also details viable fall-back options. Effective wireless coupling to hearing technologies, with the avoidance of interferences with assistive devices, and beyond, is also addressed. The necessary background information and specifications containing technical parameters is provided as informative annexes to the present document. The present document additionally provides guidance on how to implement the interoperable total conversation services for emergency communications so that the needs of all users, including people with disabilities, are met. As such the present document complements other standards developed in the area, such as EN 301 549 [1] and ETSI TS 101 470 [29].

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found in the ETSI docbox.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long-term validity.

The following referenced documents are necessary for the application of the present document.

[1] EN 301 549 (V3.2.1) (2021-03): "Accessibility requirements for ICT products and services" (jointly produced by ETSI/CEN/CENELEC).

[2] ETSI EN 301 489-1: "ElectroMagnetic Compatibility (EMC) standard for radio equipment and services; Part 1: Common technical requirements; Harmonised Standard for ElectroMagnetic Compatibility".

[3] ETSI EN 301 489-52: "ElectroMagnetic Compatibility (EMC) standard for radio equipment and services; Part 52: Specific conditions for Cellular Communication User Equipment (UE) radio and ancillary equipment; Harmonised Standard for ElectroMagnetic Compatibility".

[4] ETSI ES 202 975: "Human Factors (HF); Requirements for relay services".

[5] ETSI TS 103 479: "Emergency Communications (EMTEL); Core elements for network independent access to emergency services".

[6] ETSI TS 103 919 (V1.1.1) (2024-08): "Emergency Communications (EMTEL); Accessibility and interoperability of emergency communications and for the answering of emergency communications by the public safety answering points (PSAPs) (including to the single European Emergency number 112)".

[7] ETSI TS 124 229: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; 5G; IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3 (3GPP TS 24.229)".

[8] ETSI TS 126 114: "Universal Mobile Telecommunications System (UMTS); LTE; 5G; IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction (3GPP TS 26.114)".

[9] IEC 60118-4: "Electroacoustics - Hearing aids - Part 4: Induction-loop systems for hearing aid purposes - System performance requirements".

[10] Recommendation ITU-T G.114: "One-way transmission time".

[11] Recommendation ITU-T G.722 (2012): "7 kHz audio-coding within 64 kbit/s".

[12] Recommendation ITU-T G.722.2 (2003): "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)".

[13] Recommendation ITU-T H.264: "Advanced video coding for generic audiovisual services".

[14] ITU-T H-series supplement 1 (1999): "Application profile - Sign language and lip-reading realtime conversation using low bit-rate video communication".

[15] Recommendation ITU-T P.1305: "Effect of delays on telemeeting quality".

[16] Recommendation ITU-T T.140: "Protocol for multimedia application text conversation" (including its Addendum 1).

[17] IETF RFC 3261: "SIP: Session Initiation Protocol".

[18] IETF RFC 3550: "RTP: A Transport Protocol for Real-Time Applications", H. Schulzrinne et.al., 2003.

[19] IETF RFC 3711: "The Secure Real-time Transport Protocol (SRTP)", Baugher M., McGrew D., Naslund M., Carrara E., and Norrman K., DOI 10.17487/RFC3711, March 2004.

[20] IETF RFC 4103: "RTP Payload for Text Conversation", G. Hellstrom, P. Jones, 2005.

[21] IETF RFC 6184: "RTP Payload Format for H.264 Video", Wang Y.-K., Even R., Kristensen T., and Jesup R., 2011.

[22] IETF RFC 8445: "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal", Keranen A., Holmberg C., and Rosenberg J., July 2018.

[23] IETF RFC 8446: "The Transport Layer Security (TLS) Protocol Version 1.3", Rescorla E., August 2018.

[24] IETF RFC 8866: "SDP: Session Description Protocol", Begen A., Kyzivat P., Perkins C., and Handley M., January 2021.

[25] IETF RFC 9071: "RTP-Mixer Formatting of Multiparty Real-Time Text", 2021.

[26] IETF RFC 9110: "HTTP Semantics", 2022.

[27] IETF RFC 9147: "The Datagram Transport Layer Security (DTLS) Protocol Version 1.3", 2022.

[28] W3C® WebRTC: "Real-Time Communication in Browsers".

[29] ETSI TS 101 470: "Emergency Communications (EMTEL); Total Conversation Access to Emergency Services".

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long-term validity.

The following referenced documents may be useful in implementing an ETSI deliverable or add to the reader's understanding, but are not required for conformance to the present document.

[i.1] ATIS-0700029: "Real Time Text Mobile Device Behavior".

[i.2] ATIS-0700030: "Real Time Text End-to-End Service Description Specification".

[i.3] BEREC: "BEREC database of Access to Emergency Services in EU".

[i.4] BEREC: "Report on measures for ensuring equivalence of access and choice for disabled end-users", BEREC 2022.

[i.5] EENA: "REACH112: REsponding to All Citizens Needing Help (2009-2012)", 2012.

[i.6] EN 17161:2019: "Design for All: Accessibility" (produced by CEN/CENELEC).

[i.7] ISO 9241-20:2021: "Ergonomics of human-system interaction — Part 20: Accessibility guidelines for information/communication technology (ICT) equipment and services".

[i.8] ISO 9241-110:2020: "Ergonomics of human-system interaction — Part 110: Interaction principles".

[i.9] Ericsson® (November 2023): "Ericsson Mobility Report".

[i.10] Ericsson®: "Ericsson Mobility Visualizer".

[i.11] ETSI EG 202 320: "Human Factors (HF); Duplex Universal Speech and Text (DUST) communications".

[i.12] ETSI TR 103 201: "Emergency Communications (EMTEL); Total Conversation for emergency communications; implementation guidelines".

[i.13] ETSI TR 103 708: "Human Factors (HF); Real-Time Text (RTT) in Multiparty Conference Calling".

[i.14] ETSI TS 103 478: "Emergency Communications (EMTEL); Pan-European Mobile Emergency Application".

[i.15] ETSI TS 103 871: "Emergency Communications (EMTEL); PEMEA Real-Time Text Extension".

[i.16] ETSI TS 103 945: "Emergency Communications (EMTEL); PEMEA Audio Video Extension".

[i.17] ETSI TS 122 101: "Universal Mobile Telecommunications System (UMTS); LTE; 5G; Service aspects; Service principles (3GPP TS 22.101)".

[i.18] ETSI TS 122 173: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; IP Multimedia Core Network Subsystem (IMS) Multimedia Telephony Service and supplementary services; Stage 1 (3GPP TS 22.173)".

[i.19] ETSI TS 122 226: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Global Text Telephony (GTT); Stage 1 (3GPP TS 22.226)".

[i.20] ETSI TS 122 228: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Service requirements for the Internet Protocol (IP) multimedia core network subsystem (IMS); Stage 1 (3GPP TS 22.228)".

[i.21] ETSI TS 123 167: "Universal Mobile Telecommunications System (UMTS); LTE; IP Multimedia Subsystem (IMS) emergency sessions (3GPP TS 23.167)".

[i.22] ETSI TS 123 226: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Global text telephony (GTT); Stage 2 (3GPP TS 23.226)".

[i.23] ETSI TS 123 228: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; IP Multimedia Subsystem (IMS); Stage 2 (3GPP TS 23.228)".

[i.24] ETSI TS 124 173: "Universal Mobile Telecommunications System (UMTS); LTE; 5G; IMS Multimedia telephony communication service and supplementary services; Stage 3 (3GPP TS 24.173)".

[i.25] ETSI TS 124 238: "Universal Mobile Telecommunications System (UMTS); LTE; Session Initiation Protocol (SIP) based user configuration; Stage 3 (3GPP TS 24.238)".

[i.26] ETSI TS 126 441: "Universal Mobile Telecommunications System (UMTS); LTE; 5G; Codec for Enhanced Voice Services (EVS); General overview".

[i.27] ETSI EN 303 645 (V2.1.1) (2020-06): "CYBER; Cyber Security for Consumer Internet of Things: Baseline Requirements".

[i.28] Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services.

[i.29] FCC: "Real-Time Text: Improving Accessible Telecommunications".

[i.30] GSMA™: "GSMA Foundry 5G New Calling whitepaper 2023".

[i.31] GSMA™ IR.92: "IMS profile for Voice and SMS".

[i.32] GSMA™ IR.94: "IMS Profile for Conversational Video Service v16.0".

[i.33] GSMA™ NG.114: "IMS Profile for Voice, Video and Messaging over 5GS".

[i.34] IETF RFC 3986: "Uniform Resource Identifier (URI): Generic Syntax".

[i.35] IETF RFC 5194: "Framework for Real-Time Text over IP Using the Session Initiation Protocol (SIP)".

[i.36] IETF RFC 6116: "The E.164 to Uniform Resource Identifiers (URI). Dynamic Delegation Discovery System (DDDS) Application (ENUM)".

[i.37] IETF RFC 8831: "WebRTC Data Channels".

[i.38] IETF RFC 8832: "WebRTC Data Channel Establishment Protocol", 2021.

[i.39] IETF RFC 8864: "Negotiation Data Channels Using the Session Description Protocol (SDP)", 2021.

[i.40] IETF RFC 8865: "T.140 Real-Time Text Conversation over WebRTC Data Channels", Holmberg C. and Hellström G., 2021.

[i.41] IETF RFC 8373: "Negotiating Human Language in Real-Time Communications".

[i.42] IETF RFC 9248: "Interoperability Profile for Relay User Equipment".

[i.43] IETF Secure Telephone Identity Revisited (STIR).

[i.44] Recommendation ITU-T F.700: "Framework Recommendation for multimedia services".

[i.45] Recommendation ITU-T F.703: "Multimedia conversational services".

[i.46] Recommendation ITU-T F.930: "Multimedia telecommunication relay services".

[i.47] Recommendation ITU-T H.265: "High efficiency video coding".

[i.48] Recommendation ITU-T V.18: "Operational and Interworking for DCEs operating in the text telephone mode".

[i.49] D. Lewin, B. Glennon and B. Hoemburg: "Voice telephony services for deaf people. An independent report for Ofcom", Plum Consulting, 2009.

[i.50] NENA STA-010.3: "NENA i3 Standard for Next Generation 9-1-1".

[i.51] NENA: "NENA Video Relay Service & IP Relay Service PSAP Interaction Information Document".

[i.52] Nielsen T.: "Implementation of RTT and Total Conversation in Europe", EENA, 1 March 2023.

[i.53] OECD broadband statistics: "Top 10 countries in mobile date usage per mobile broadband subscription", OECD, June 2024.

NOTE: Available at https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/broadband-statistics/data/1-14-mobile-data-usage-top-10.xls.

[i.54] D. S. Ray and E. J. Ray: "Unix and Linux: Visual QuickStart Guide", Peachpit Press, 1998.

[i.55] The OpenVMS PHONE Facility.

[i.56] Tudor B.: "Why video-calling hasn't made the telecoms companies rich", The Guardian, 10.08.2007.

[i.57] W3C® Working Group Note 25 May 2021: "RTC Accessibility User Requirements".

[i.58] "History of videotelephony" from Wikipedia®.

[i.59] Wytec: "Executive Summary 2023".

[i.60] Digital In the Round: "Video conferencing statistics", July 10, 2021.

[i.61] ETSI ES 201 275 (V1.1.1) (1998-08): "Human Factors (HF); User control procedures in basic call, point-to-point connections, for Integrated Services Digital Network (ISDN) videotelephony".

[i.62] ETSI EG 202 116 (V1.2.2) (2009-03): "Human Factors (HF); Guidelines for ICT products and services; "Design for All"".

[i.63] ETSI TS 126 235: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Packet switched conversational multimedia applications; Default codecs (3GPP TS 26.235)".

[i.64] ISO/IEC 10646: "Information technology — Universal coded character set (UCS)".

[i.65] Michael Kalloniatis, Charles Luu: "Visual Acuity", in Webvision, The Organization of the Retina and Visual System, Part VIII.

[i.66] IETF RFC 8862: "Best Practices for Securing RTP Media Signaled with SIP".

[i.67] Centre of Expertise for Accessible Client Service (2021): "Accessibility Playbook - Delivering accessible client service".

[i.68] Accessibility for Ontarians with Disabilities Act, 2005.

[i.69] Drullman R., Festen J. M., & Plomp R. (1994): "Effect of reducing slow temporal modulations on speech reception", The Journal of the Acoustical Society of America, 95(5), 2671-2680.

[i.70] N.R. French & J.C. Steinberg: "Factors Governing the Intelligibility of Speech Sounds", JASA vol. 19, No 1, 1947.

[i.71] Kozma-Spytek L., Tucker P., Vogler C.: "Audio-visual speech understanding in simulated telephony applications by individuals with hearing loss", Proc 15th Int ACM SIGACCESS Conf Comput Access (ASSETS 2013), 2013. Bellevue, WA, United States: Association for Computing Machinery. doi: 10.1145/2513383.2517032.

[i.72] Brysbaert, M., 2019: "How many words do we read per minute? a review and meta-analysis of reading rate", Journal Of Memory And Language 109.

[i.73] IEC 60118-13: "Electroacoustics - Hearing aids - Part 13: Requirements and methods of measurement for electromagnetic immunity to mobile digital wireless devices".

[i.74] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[i.75] Recommendation ITU-T H.320: "Narrow-band visual telephone systems and terminal equipment".

[i.76] Recommendation ITU-T H.324: "Terminal for low bit-rate multimedia communication".

[i.77] Recommendation ITU-T H.323: "Packet-based multimedia communications systems".

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the following terms apply:

Assistive Technology (AT): any item, piece of equipment, service or product system including software that is used to increase, maintain, substitute or improve functional capabilities of persons with disabilities or for, alleviation and compensation of impairments, activity limitations or participation restrictions

conversation: any two-way human language communication in real-time between people or between people and ICT

NOTE: Human languages take different forms including spoken, written or signed.

emergency communication: communication by means of interpersonal communications services between an end-user and the PSAP with the goal to request and receive emergency relief from emergency services

media: form of transmitting and presenting information to the user of electronic communications

NOTE: Electronic communication media include text, video and audio (voice).

modality: particular way in which communication is experienced or is expressed

NOTE: The most valid examples for the present document are signed (= using sign language), written and spoken modalities.

multimedia: combinations of static and/or dynamic media presented simultaneously

NOTE: Examples of multimedia include combinations of text and video, or audio and animation, etc.

pre-registration: registration of identifier and other characteristics of a user of a specific service, or a specific feature available via a service, with the service provider before use of the service or feature is requested

NOTE 1: Examples are pre-registration for use of total conversation with language and modality preferences in emergency communications, pre-registration for use of a relay service interoperating with the total conversation service.

NOTE 2: Examples of registrations not regarded to be pre-registrations are: signing up for a communication service providing total conversation for all users, and making settings in the user equipment about language and modality preferences for communication.

Public Safety Answering Point (PSAP): Public Safety Answering Point with call takers handling the emergency communications

Real-Time Text (RTT): form of text conversation in point to point situations or in multipoint conferencing where the text being entered is sent in such a way that the communication is perceived by the user as being continuous on a character-by-character basis

total conversation: bidirectional symmetric real-time transfer of motion video, real-time text and voice between users in two or more locations

total conversation service: multimedia real time conversation service that provides bidirectional symmetric real time transfer of motion video, real-time text and voice between users in two or more locations

user equipment: combined hardware and software used by a user

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

3GPP	Third Generation Partnership Project
AAC	Augmentative and Alternative Communication
AI	Artificial Intelligence
ALD	Assistive Listening Devices
ATIS	Alliance for Telecommunications Industry Solutions
CIA	Confidentiality Integrity Availability
EAA	European Accessibility Act
ES	ETSI Standard
GDPR	General Data Protection Regulation
GSMA	Global System for Mobile Communication
HAC	Hearing Aid Compatibility
ICT	Information and Communication Technology
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem
IVR	Interactive Voice Response
MFA	Multi-factor Authentication
MITM	Man-In-The-Middle
MTSI	Mobile Telephony Service for IMS
PSAP	Public Safety Answering Point
PSTN	Public Switched Telephone Network
QVGA	Quarter Video Graphics Array
RF	Radio Frequency
RTP	Real-time Transport Protocol
RTT	Real-Time Text
SBC	Session Border Control
SIP	Session Initiation Protocol
SRTP	Secure Real-time Transfer Protocol
SSO	Single Sign-on
STIR	Secure Telephone Identity Revisited
STT	Speech To Text
TLS	Transport Layer Security
TTS	Text To Speech
UE	User Equipment
UICC	Universal Integrated Circuit Card
VoIP	Voice over IP
WebRTC	Web Real-Time Communication
wpm	words per minute

4 Total conversation service: General characteristics and usage scenarios

4.1 General functionality of total conversation services

4.1.1 Type of service

A total conversation service is an electronic communication service which provides real-time conversational communication between two or more participants with total conversation User Equipment (UE), using real-time text, video and voice according to the preferences of each user.

Annex A of the present document provides a background. Annex B provides a commented bibliography.

NOTE: Total conversation participants may be automata or humans.

4.1.2 Provided media

4.1.2.1 General about provided media

A total conversation service **shall** be capable of simultaneously providing real-time text, video, and voice with a quality that meets the accessibility needs and preferences of individuals. The minimum quality and performance requirements and the requirement for user preferences are specified in clauses 5.2 and 6.2.4 of the present document, respectively, of the present document. The actual inclusion of these media in a communication may vary depending on the actions taken by the users or the capabilities of individual end user devices.

4.1.2.2 Real-time text

Real-time text is a conversational medium sent while it is produced character by character or in small groups of rapidly entered characters. Real-time text provides the opportunity to the receiver(s) of text to follow what the sender expresses in text without waiting. This is of benefit to the receivers by establishing a good feeling of contact in sessions and by allowing them to start to respond quickly.

The use of real-time text can be for carrying the main part of the information in the session, or for just one direction of the session while speech is used in the other direction, or for occasional use for explanations in a session dominated by other media.

4.1.2.3 Video

Video in a total conversation session can be used for varying purposes. It can be used for showing the participants in a session in order to enhance the feeling of contact between them. It can be used for showing objects and contextual views during a session.

NOTE: Contextual views may be a view of surroundings, objects that are talked about, or similar. May be especially useful to show an emergency scene.

The main accessibility benefit of video in interpersonal communications services is to allow the use of sign language communication for rapid and convenient interaction between deaf persons. Video also enhances the opportunities for hard of hearing persons to better understand spoken language when the talking person's lips can be seen.

4.1.2.4 Voice

The audio component in total conversation can be used for speech communication as in any other interpersonal communication service. It can also be used in various combinations with the other two real-time media in the session.

4.1.2.5 Other components in the session

The three main media in a total conversation session can be combined with other communication components just as in other interpersonal communications services. Examples are:

- Presentation of documents, e.g. for presentation in a conference.
- Message based text chat channels/rooms as commonly used in video conferencing for documented interactions (these serve a different purpose than RTT)".
- File transfers, e.g. to transfer files discussed in a meeting.
- Hand raising and other reactions, e.g. for managing meetings in all media in uniform ways.
- Turn indication, for management of larger meetings where strict order of contributors is needed.

4.2 Usage scenarios: Typical communication situations

4.2.1 Communication with any other users

4.2.1.1 General

Total conversation is typically used in interpersonal communications sessions with two or more participants. The participants use the conversational media for any of the usage types of the provided media described in clause 4.1.2 of the present document.

In addition to allowing the users to freely use available communication media, it **should** also be possible for the users to communicate across the various user equipment types and different interoperating services.

4.2.1.2 Multimodal communication

Typically, the users have competence in the same communication modality and are using that modality predominantly, while occasionally using the other modalities as complements. Examples are:

- Video & RTT both ways: Users of sign language using sign language, and occasionally using real-time text for numbers or terms where spelling is important or when it is convenient to have the transmitted items stationary on the screen.
- RTT & video both ways: Users of real-time text using real-time text as main communication modality, and also seeing each other.
- Audio, video & RTT both ways: Speech users mainly using voice, also seeing each other and having real-time text for rapidly complementing the communication with numbers or terms that require exact spelling.
- RTT & video both ways: A user of sign language initiates a session with the intention of having a sign language conversation with a friend, but another person without sign language competence answers. They then divert to using real-time text as a common conversational modality.

There are also situations where different modalities are used with different intensity by different participants of the conversation. The following examples illustrate how other modalities may be important even when speech is used by all participants of the conversation:

- Speech both ways, RTT captioning included to one user by automatic or manual speech-to-text service to facilitate/enhance the understanding of the spoken information.
- Speech both ways, complemented by video of the speaking persons for lip-reading to improve the understanding.

The additional media: RTT and/or video, are useful for speaking hard of hearing users, non-native speakers, and persons with cognitive disabilities.

The total conversation service **shall** be capable of conveying the media from each participant in a session to all other participants in the session regardless of the communication modalities used by individual participants. With total conversation different participants may use different communication modalities. An example may be one user speaking and the other only sending RTT. This may be the case of participants with speech disabilities or those who are deaf or hard of hearing, but may also be true for people who cannot speak due to their particular situation at the moment (e.g. a noisy environment, or where they do not wish to be overheard).

4.2.1.3 Multiparty communication

The most basic implementations of total conversation all allow simple person-to-person communication. But to be fully effective, it is essential for total conversation to also allow conversations between multiple users. The functionality available to users in multiparty total conversation **shall** be the same as the functionality available to users in multiparty voice communication on the underlying communication platform.

In traditional phone networks the number of people who can participate in one call is usually limited, often to three or four people, and it **should** be possible for total conversation to take place between the same number of participants. For communications platforms that support conferencing between many people, the full total conversation experience **should** also be available to all those participants.

Multiparty total conversation is especially valuable in cases where a relay service is supporting the communication needs of one of the participants in the conversation (clause 4.2.2 of the present document goes into more depth regarding relay services). This valuable capability becomes essential in several emergency communication scenarios where the involvement of a relay service or another assisting service is required to ensure that the emergency communication is effectively handled. Whether multiparty communication is supported or not could be a matter of life or death. Emergency service usage is covered in more depth in clause 4.2.3 of the present document.

Some of the performance related issues in supporting multiparty communication can be found in clause 5.2.5 of the present document.

4.2.1.4 Communication across different services

Ideally total conversation communications **should** be provided in the same universal way as for traditional, number based voice communication. Any user **should** be able to reach anyone else with total conversation, regardless of whether they use the same or different provider for their total conversation service, or whether they use the same or different kinds of user equipment.

While the connectivity is supported between different user equipment, the total conversation services available when the present document was authored rarely provide the cross-service interoperability. Such interoperability is technologically achievable as explained in clauses 5.3 and 5.4 of the present document and would be beneficial for users.

4.2.2 Communication supported by relay services and transcription functions

Modality conversion is valuable for accessibility to interpersonal communications. Relay services and transcription functions perform conversion or interpretation between different forms of communication. Total conversation is very suitable for access to and use of relay services and transcription functions.

Forms of relay services are:

- Video relay service interpreting between sign language in video and speech in audio.
- Text relay service converting between real-time text and speech.
- Captioned telephony adding real-time text transcription (subtitles) to voice communication.
- Speech to speech service supporting persons with speech related disabilities and cognitive disabilities.

Requirements on relay services are provided in ETSI ES 202 975 [4] and Recommendation ITU-T F.930 [i.46].

Transcription functions may be also used by the translation services which not only convert the speech/text to text/speech but also translate it into a different human language. Considerations for translation services are detailed in clause 6.4.3.3 of the present document.

Speech can be automatically transcribed to real-time text with short delay, often under a second.

Speech can also be transcribed to real-time text manually, resulting in longer transcription delays.

Synchronization of speech and subtitling is further discussed in clause 5.2.4.3 of the present document.

4.2.3 Using total conversation for emergency communication

This clause applies where a total conversation service is required to provide or provides, emergency communications.

It **shall** be possible for total conversation users in emergency to initiate emergency communication and communicate in a modality and language the user in emergency can handle.

It **shall** also be possible to freely choose the desired mix of media activated, even if the actual human communication is ongoing only in one modality. For example, if the user is able to only convey and receive the information using RTT, availability of audio and video in addition is still desirable to provide the emergency call-takers with more information; additionally, the view of the professional call-taker has a calming effect on the user in emergency. The detailed specifications about emergency communication initialization and routing are provided in the emergency communication specific standard ETSI TS 103 919 [6].

NOTE: ETSI TS 103 919 [6] forms the base for work in progress when the present document was authored, with a draft harmonised standard on accessible emergency communications with the same title.

4.2.4 Usage with assistive technologies

4.2.4.1 General on assistive technologies

The basic user equipment commonly used for total conversation is usually not sufficiently accessible for convenient use by all potential users. Some users need to use assistive technologies to operate user equipment, and they will need to be able to handle the total conversation management and communication using those assistive technologies.

Assistive technology may be additional devices with software, or just software, or services to be understood as documented platform accessibility services as defined in EN 301 549 [1]. Some user equipment allows the attachment of external assistive devices to allow them to use the equipment. Other user equipment may not allow a user's assistive devices to be connected or allow connected assistive devices to access all aspects of equipment that the user needs to operate. Functionality that cannot be accessed from assistive devices or from installed assistive software is referred to as "closed functionality". Mobile telephones rarely allow full external access by assistive technologies, but they usually have inbuilt capabilities that replace the need to attach or install such assistive technologies for some, but not all, users.

Depending on the communication modality, different assistive technologies may be relevant. A review of some of the relevant technologies is provided in the remainder of this clause. Clause 5.5.2 of the present document contains the requirements for total conversation user equipment to ensure compatibility with relevant assistive technologies.

4.2.4.2 Accessibility of alerts in communication initialization

Assistive technologies supporting alerting on incoming communication include pocket, wrist or other types of devices. They are wirelessly activated from the main user equipment and provide haptic, strong sound and/or visual alerts to notify persons with limited or no hearing about incoming requests for communication.

4.2.4.3 Accessibility of voice communication by assistive technologies

Assistive technologies for voice input consist of:

- Augmentative and Alternative Communication (AAC) devices and tools that help individuals with speech impairments to communicate, including devices that converts speech to text or symbols.
- Emerging sign-to-speech technologies aiming to provide automated translation of sign language to text or voice. Although there are some existing solutions, a wider availability is still limited due to the intrinsic complexities of sign languages.

Assistive technologies for voice output/reception consist of:

- Hearing aids and cochlear implants - devices that amplify sound or directly stimulate the auditory nerve to assist individuals with hearing impairments in perceiving sound.
- Assistive Listening Devices (ALDs) that enhance the sound in specific environments, making it easier for individuals to hear.
- Captioning and subtitling tools that are based on speech recognition technologies, converting spoken language into text.
- Emerging speech-to-sign technologies: Research on systems that use machine/AI-assisted automated speech to sign language generation. However, these systems still face challenges due to the intrinsic complexities of sign languages.

4.2.4.4 Accessibility of real-time text communication by assistive technologies

Assistive technologies for text handling in communication by assistive technologies are provided in the following forms:

- Assistive technologies for text input:
  - Devices that allow users with physical disabilities to input text and control the interaction with ICT devices without using a traditional keyboard or mouse.
  - Braille keyboards, devices that support Braille users.
  - Software that converts spoken language into text, allowing users to control devices and input text via voice dictation.
  - Adapted mechanisms for navigation and operation support and text entry in control and communication with RTT in total conversation communications. These mechanisms include eye gaze control, track-ball control and suck-and-blow control, etc.
- Assistive technologies for text output:
  - Screen magnifying software (or hardware) that enlarges text and images on a screen, making it easier for individuals with low vision to see content.
  - Braille displays, devices that support Braille users.
  - Screen reader software reading out the user interface contents, enabling users with vision limitations to operate the user interface and to receive the content of the RTT communication in total conversation.
  - Text To Speech (TTS) tools can be used to read aloud RTT communication.

NOTE: In the context of total conversation, it is screen readers that are most relevant for users with disabilities. TTS tools which offer a simple conversion of received written text into spoken words, and do not support the overall user interaction, are likely to be used occasionally, when needed, by users who do not need screen readers. For example, TTS, rather than a screen reader, can be used to hear the RTT communication when driving.

- Screen reader software that outputs to a Braille display which enables users with vision limitations, e.g. persons with deafblindness, to read the content of and interact with the user interface to enable user interface interaction for both operation and RTT transmission in total conversation.
- Hearing devices with wireless connection to the main user equipment for access to audio during total conversation by persons with hearing limitations. These come in various forms, e.g. with wireless connection directly to hearing aids, wireless connection to control unit for various enhanced hearing devices, wired or wireless connection to inductive loop devices for wireless connection to hearing aids.
- Software or service for automatic speech-to-text transcoding, supporting hearing-impaired or deaf users.

5 Requirements for total conversation service provision

5.1 General

Provision of total conversation services has many similarities with provision of video telephony and video conferencing. The difference is that the real-time text media is included together with video and audio and that there are accessibility motivated quality requirements on the media performance. The present clause 5 provides performance requirements on the media and presents a conceptual system architecture for interoperability and suitable communication protocols. It also outlines the requirements for user equipment, including the requirements for interoperability with assistive technologies.

5.2 Performance requirements on media

5.2.1 Real-time text

5.2.1.1 End-to-end delay and smoothness of presentation

The main performance expectation on real-time text is that each received text character appear with minimal delay from when the character was entered by a sending user. According to Recommendation ITU-T F.703 [i.45] one second end-to-end delay is the maximum acceptable in an interpersonal communication situation.

This is longer than the accepted delay for video and audio. There is a natural background for that longer accepted delay in that when text is produced through a keyboard, there may be natural delays introduced by a typing person hesitating on spelling or not finding the key or any other physical reason.

Longer delays than one second cause annoyance and uncertainty if the party who is expected to send text at the moment has understood the previous entry or has lost focus on the conversation. A longer delay also causes risk of turn collisions when one party starts restating a question at the same time as the answer from another party reaches the destination.

Shorter delay than one second has not been observed to contribute much to conversation satisfaction. There may however be specific situations, such as in multiparty conferences or gaming where a shorter delay can be found important or active floor-control mechanisms are required. For more details see clause 5.2.5 of the present document.

Characters are commonly grouped in RTT transmission for transmission efficiency in groups of characters being entered during the same short transmission interval. The interval **shall** be at most 500 ms when there is any new text to transmit, see clause 6.2.4 of EN 301 549 [1]. A 300 ms interval **should** be used, giving a good compromise between smooth presentation good responsivity and transmission efficiency.

For even smoother playout of received characters, the receiving user equipment **may** apply a delay of 30 ms between presentation of characters from the same source being just above a commonly achieved reading speed of users to be between 200 and 300 words per minute (with a word being in average 5 characters including separators). See Brysbaert [i.72].

NOTE: It is advisable to provide functionality to control the speed of text presentation (also subtitling), allowing the users to adjust the speed to their individual preferences in a given communication situation.

5.2.1.2 Reliability

All electronic communications are done with a risk for loss or corruption of data. Retransmissions are used to reduce the risk of loss or corruption. The better protection against loss and corruption there is, the longer time it takes to regain normal communication in case of transmission errors.

RTT transmission **shall** be protected against character loss and corruption by network transmission errors. A balance between remaining error rate and rapid recovery of real-time performance **shall** be maintained. A requirement stated in Recommendations ITU-T F.700 [i.44] and F.703 [i.45] is that no more than one character out of 500 is lost or corrupted in network conditions where voice quality is severely affected but still can be regarded as useable. The intention of setting this allowable error rate is that the error rate from communication **should** be lower than the errors normally appearing during human input of text.

If unrecoverable errors appear, their place in the resulting text **should** be marked. If the transmission is momentarily blocked because of retransmissions, then such blockages **shall** be no longer than 15 seconds before the error is regarded unrecoverable.

5.2.1.3 Character representation and editing

It **shall** be possible to transmit and present characters in the International character set Unicode, including emojis.

Brief editing of text **shall** be possible, including new lines and erasing latest entered text (also including erasing new lines).

5.2.1.4 Other performance related aspects of RTT

Requirements on many other aspects related to performance of real-time text entry, transmission and presentation can be found in EN 301 549 [1] clause 6.2, and in ETSI TR 103 708 [i.13].

5.2.2 Video

5.2.2.1 End-to-end delay of video

The delay of video from capturing in a camera to presentation on a screen by the receiver **shall** be at most 400 ms both for use for viewing a talking participant and for viewing a participant using sign language. See ITU-T H-series Supplement 1 [14] and Recommendation ITU-T P.1305 [15]. Higher satisfaction is observed at shorter delays down to 150 ms.

5.2.2.2 Smoothness of motion reproduction

The smoothness of motion reproduction needed for good perception can be expressed in frames per second. For both viewing a talking person and a person using sign language, at least 20 frames per second **shall** be communicated and presented. See ITU-T H-series Supplement 1 [14].

NOTE: Most current bidirectional video communication applications provide a nominal rate of 30 frames per second but reduce the rate when large and heavy motion is visible. This makes it often feasible to achieve the required rate of 20 fps, but it should be observed that there usually is a balance between spatial and temporal resolution, so that an unnecessarily high spatial resolution setting can result in a frame rate lower than required.

5.2.2.3 Spatial resolution and sharpness

The spatial resolution (= sharpness) of video **should** be sufficient for perception of small details of importance for understanding sign language (like fingers or gaze direction) and for perception of facial expressions for lip-reading. With a view of one person from skull to waist, the spatial resolution **shall** correspond at least to QVGA (320x240 pixels). See ITU-T H-series Supplement 1 [14].

Reduction of spatial resolution is not only caused by the compression, coding and transmission of video. It is also caused by the characteristics of the camera. The camera used **should** be able to produce images during heavy motion of a part of the image corresponding to moving arms, hands and face of one person, which may be about 25 % of the image. Camera performance is often heavily depending on light conditions, so for all these aspects, the camera **shall** under intended working conditions provide images which are not more blurred than what the QVGA spatial resolution causes.

NOTE: This spatial resolution requirement is usually achieved and surpassed in most current bidirectional video communication implementations used by the general communications users.

5.2.2.4 Evaluation of video quality

The video quality requirements presented above are usually achieved and surpassed by modern equipment and services. There is however a risk that spatial resolution is over-emphasized in video settings at the expense of smooth and detailed representation of motion. A verification of achieved video quality for sign language and lip-reading can therefore be motivated. See Annex C for a simple video quality evaluation method.

5.2.3 Voice

5.2.3.1 End to end delay for voice communication

The delay of voice from capturing in a microphone to presentation by the receiver **shall** not be more than 400 ms. See Recommendations ITU-T G.114 [10] and ITU-T P.1305 [15]. Higher satisfaction is observed at shorter delays down to 150 ms.

5.2.3.2 Frequency range of voice communication

The most important frequency range for speech perception is 250 Hz to 7 000 Hz. See French [i.70], Figure 12. Frequencies between 100 Hz and 250 Hz can contribute to person recognition and other factors of additional value. See Drullman [i.69],

The traditional frequency range of telephony of 300 Hz to 3 400 Hz is not sufficient for ease of perception of speech. Wide band audio defined to be at least up to 7 000 Hz **shall** be provided for good opportunities of speech perception.

Even if low frequencies from 85 Hz occur in human voices, the voices are rich of formants, and therefore it is usually sufficient to reproduce sound from 250 Hz and up for good speech perception. Frequencies from 250 Hz **shall** be captured, transmitted and be possible to reproduce by a total conversation service. For the requirement on user equipment see clause 5.5.4 of the present document.

Frequencies between 100 Hz and 250 Hz **may** add value for person and emotion recognition. Clause 5.5.4 of the present document requires the user equipment to capture these frequencies, and the total conversation service **shall** convey them. The actual presentation to a receiving user of these frequencies may require use of attached listening devices, see clause 5.5.4 of the present document.

5.2.4 Synchronization of multiple communication media

5.2.4.1 General

Synchronization of presentation of different media in a total conversation session is important for understandability of the communication. The need for synchronization is essential when one person's communication is transmitted with more than one media: a video of speaking person (voice and video synchronization), or subtitling of video of speaking person (text and video synchronization).

5.2.4.2 Voice and video synchronization

Any total conversation implementation **shall** ensure appropriate synchronization of audio and video. This requirement provides for usability of the combination of audio and video for lip-reading, an also for a general satisfaction of a view of lip synch.:

- Video **shall** not be presented more than 100 ms before voice.
- Video **shall** not be presented more than 100 ms after voice.

This is under the condition that the frame rate of video meets its requirements as specified in clause 5.2.2 of the present document. See Kozma-Spytek et. al [i.71], section 3.3.4.2 and Table 4.

5.2.4.3 Subtitling synchronization

Whenever subtitling for spoken input is provided, and real-time subtitles are available (see clause 4.2.2 of the present document), they **shall** be presented as promptly as possible following the voice input.

Automated transcoding of speech or sign language to RTT, the delay of subtitles in relation to the speech **should** not exceed 2 seconds to ensure usability and effective comprehension.

NOTE 1: In case of automatic transcoding of speech or sign language to RTT, the transcoding process usually need to hear the full word before it can be written. Average conversational speed of human speech in English language is 2 words per second (wps) or 120 words per minute (wpm). That causes an initial mean delay of at least 500 ms before the word is started to be converted into the RTT. Considering the maximum allowed RTT end-to-end delay of 1 second, and the approximate delay for voice (of about 200 ms in transmission), the automatically captioned RTT can be delayed by 1,3 seconds after voice. In reality, the delay is likely to be shorter because both the transmission and buffering delays for RTT are likely to be shorter than the assumed 500 ms, and rather equal to 300 ms for transmission and 150 ms for buffering (see clause 5.2.1 of the present document), resulting in the overall synchronization delay of the order of magnitude of 750 ms (500 ms - 200 ms + 300 ms + 150 ms), rather than 1,3 seconds for RTT after voice. This results in a synchronization within the acceptable 2 seconds delay of subtitles after voice as described in clause 4.2.2 of the present document.

NOTE 2: The subtitling delay will be impacted also by audio quality.

For human generated transcoding, the delay however is usually longer.

For two-party use, the time for production of text can be allowed to be longer if the speaking and hearing user are informed that a relay service is in use.

In multiparty and multimodal conversation, the conversation needs to be managed to allow users relaying on captions for effective participations as specified in clause 5.2.5.3 of the present document.

5.2.5 Multiparty considerations

5.2.5.1 Video

Regarding video in multiparty communications, the greatest challenge concerns presentation of communication participants. The presentation is usually combined in a group of rectangles, smaller the higher number of communication participants that get their video presented. Each presented participant takes relatively big amount of communication resources in form of bandwidth use in the network and coding and decoding processing power in the involved user equipment and multiparty bridges. Screens also have limited resolution, so each participant **shall** be presented with at least the minimum acceptable resolution as specified in clause 5.2.2 of the present document.

This may lead to limitations in how many participants can be presented with good video quality sufficient for sign language and lip reading.

Requirements for user-control of the interface are provided in clause 6.5 of the present document.

5.2.5.2 Voice

For voice, the presentation is usually mixed from multiple participants into one media stream. There is a practical limit of how many people can have an effective and meaningful unmanaged communication. The requirements for communications facilitation are contained in clause 6.4.4 of the present document.

5.2.5.3 Real-time text

The performance requirements on real-time text **shall** be fulfilled during multiparty communication.

In many configurations, the transmission of multiparty multimedia communication is co-ordinated by mixing conference bridges in the communication service. Both the service and the user equipment **shall** have the capacity to transmit, receive and present RTT from at least 3 sources sending at a rate of at least 30 characters per second for a total of at least 90 characters per second. See EN 301 549 [1].

Short delay is critical, especially in an multiparty situation. Recommendation ITU-T P.1305 [15] describes that turn-taking in unmanaged voice conversations take place dominantly with zero to 200 ms gap between the parties. The one second delay assumed as acceptable for RTT is therefore not sufficient for a real-time text user in multiparty communication. Both in a mixed speech RTT multiparty conversations and when more than three participants contribute intensively with RTT text simultaneously, the communication will be hard to follow. Therefore, it is critical to have floor control for the case of multiparty conversations with RTT participants and with participants who rely on subtitles (see clause 5.2.4.3 of the present document). The requirements for communication facilitation are presented in clause 6.4.4 of the present document.

5.3 Service provisioning

5.3.1 System architecture (informative)

5.3.1.1 Functional elements

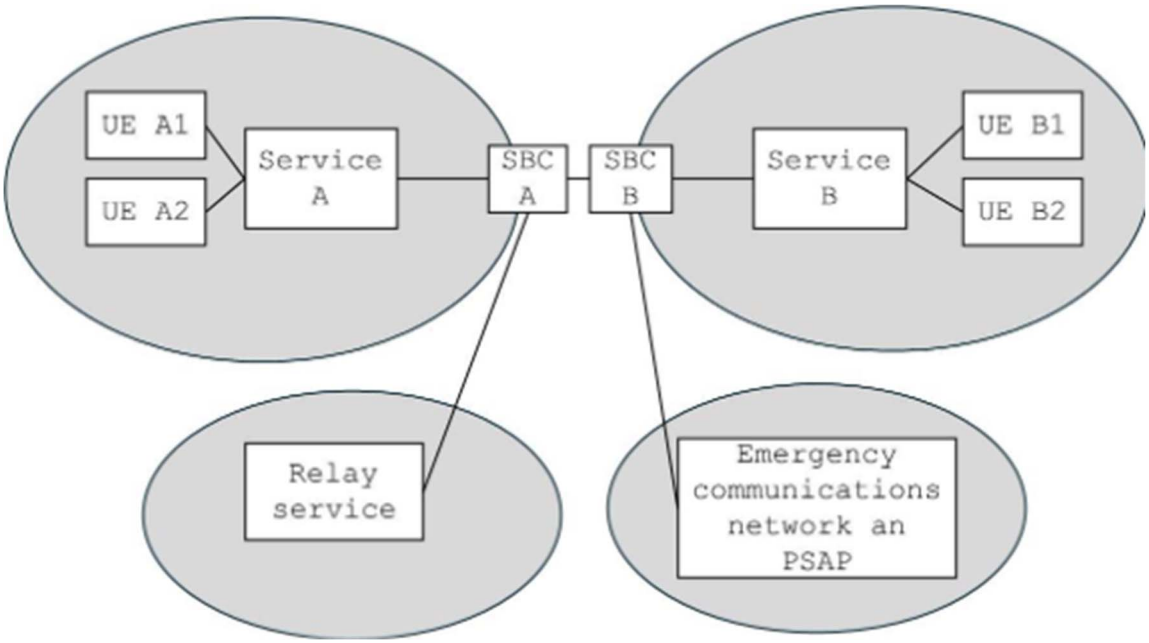

Figure 5.1: Conceptual view of a total conversation service architecture

A conceptual system architecture is provided here for the discussion of interoperability. Figure 5.1 shows the session and media handling systems and the user equipment for two communication services providing total conversation. It also shows how assisting services and emergency communication are linked to the total conversation service. The networks use packet switched communication. The lines between the functional elements are therefore symbolic representing routes where packets flow between the functional elements.

The following functional elements are shown.

Service A: A communication service providing total conversation for its users and a possibility for them to have communication with users of Service B and have emergency communications and get support by relay service for modality conversion when needed. Service A also supports multiparty communications for its users or a combination of participants from service A, Service B and the assisting services shown.

Service B: A communication service with similar offerings as Service A and with interoperability agreed and established with Service A.

UE A1: User equipment with software using the services of service A, consisting of an ICT device with screen, keyboard, camera and microphone, and equipped with software sufficient either for having total conversation communications in Service A or for fetching a web page with the communication program needed for the moment a total conversation session is started in Service A.

UE A: Similar to UE A1.

UE B1: User equipment with software similar to UE A1 but using the services of service B.

UE B2: Similar to UE B1.

Session Border Controller A (SBC A): A function that enables and checks communication external to Service A, and activates protocol conversion where needed to achieve interoperability in session control and/or media.

Session Border Controller B (SBC B): Similar to SBC A but for Service B.

Relay service available through interoperability with the services A and B for modality conversion, e.g. between sign language and speech, real-time-text and speech, speech that is hard to understand because of articulation variations and clear speech, as well as memory assisting service.

Emergency Communication Network: The network surrounding the Public Safety Answering Points and has functionality for routing emergency communications to the most appropriate PSAP.

5.3.1.2 Typical communication actions

5.3.1.2.1 Communication between user equipment within a service

When a user of a user equipment UE A1 wants to have communication with a user of another user equipment UE A2 in the same service, the user requests the UE A1 to send a request for communication to the service and includes an identifier of the other UE A2 in a form that the service recognizes and can resolve to a network address where the other UE A2 resides.

In the communication initiation request, information about the media supported by the initiating UE A1 and wanted in the communication is usually included. The service conveys this information to the target UE A2. This results usually first in an alerting by UE A2 to its user, who accepts the communication request. This causes a response from UE A2, resulting in a completed negotiation about what media to include and how these media will be coded and transported between UE A1 and UE A2.

After that the media starts to flow by the UE A1 and UE A2 sending coded media as a packet stream along paths agreed on during the initiation of the communication. The media does not need to go through any central server of the service. It is possible for the UEs to agree on sending the media directly between them. Because of this negotiation of media and coding and transport mechanisms for media, interoperability is achieved by both sides starting software codec modules corresponding to the agreed media and coding of media, and initiate transport of media with parameters matching the agreements made in the negotiation.

During the communication, the parties can send requests for changes of the communication, e.g. for adding or dropping media or for including more communications participants.

5.3.1.2.2 Communication between user equipment of different services

When a user of a user equipment UE A1 registered in one service wants to have communication with a user of another service, this can be accomplished if the services have interoperability agreements, and they have arranged provisions for the interoperability.

It is common that the services arrange for having Session Border Controllers (SBCs) through which communications with users of other services are routed, as a security precaution and for solution of differences in media coding capabilities and ways to establish communication.

The initiating user of UE A1 initiates the communication by including the phone number or identifier of the target user equipment UE B1in the initiation, and this identifier is converted to an address within service B. The resolution to an address, routable to UE B1 usually takes place in Service B. The negotiation can then take place, and it goes through the Session Border Controllers. They can, if needed, convert between different protocols for session control and for media coding and transport.

After the communication is established, media can flow as packet streams between the UE A1 and UE B1 using the coding and other parameters agreed, and usually also transported through the SBCs for any security verification and coding conversion.

5.3.1.2.3 Communication between user equipment of different technologies

The case above explains a common system architecture for cases when the communicating services implement the same or similar technologies, that the Session Border Controllers can overcome by minor conversions in session establishments or media coding. For cases when the services are based on more different technologies, it is more common to let functional elements called gateways handle the conversion. This is not shown in the architecture Figure 5.1.

5.3.1.2.4 Addressing

An important part of establishing communication between user equipment in different services is to handle the addressing of the communication to the target user equipment. Number based communication services look up the number of the target user or user equipment in a number data base and get it converted to an address, usually in the form of user@domain that can be used for finding the way to the server handling the target user. When the target user is in another service, the communication will be routed via the Session Border Controller (SBC).

Number independent services often make use of a username for addressing, with its domain assumed by default, or an explicit address based on a structure like user@domain. Protocols for the communication establishment can also be included in the address and help resolving the address to a path to take through the networks.

5.3.1.2.5 Communication with relay service support

Even if the opportunities to find common possibilities for communication increases by having the three media of total conversation available, there are many situations when the communication participants do not feel comfortable with communication in any of the common modalities. For such cases a relay service is needed in the communication.

It has been common to provide relay services only for number-based communications. The need for relay service support is equally important in number-independent communications.

Once a communication involving a relay service is established, the relay service acts on the contents in the media and converts between modalities, e.g. from sign language to speech, from speech to sign language, from real-time text to speech or from speech to real-time text. The relay service enabled in a multiparty fashion and can act in one or both communications directions in the communication.

The relay services is external to the service of the user equipment and the communication therefore is passing Session Border Controllers where protocol conversion can take place if needed.

Requirements on relay services and the access of relay services are specified in ETSI ES 202 975 [4] and EN 301 549 [1], clause 13.1.

NOTE: To ensure accessibility, using a relay service should be as simple as starting a total conversation or voice communication. The implementation of this principle primarily relies on the relay service invocation system but also requires interaction with the primary relay service user and active deployment of user profiles as described in clause 6.2.4 of the present document.

5.3.1.2.6 Aspects of total conversation user equipment important for emergency communications

This clause applies where a total conversation service is required to provide or provides, emergency communications.

Clause 9.4 of ETSI TS 103 919 [6] specifies how a total conversation user in emergency is enabled to initiate emergency communications and reach the most appropriate PSAP for handling the emergency communicating in the preferred language and modality of the user. See also clauses 4.2.3 and 6.2.4 of the present document.

The initial establishment of the emergency communication from a communications service is very similar to establishment of a communication with a user equipment of another service. A difference is that a specific identifier is used for the destination of the communication, called a "service urn".

NOTE: An example is "urn:service:sos". If the user calls "112", the user equipment initiates communication to "urn:service:sos" and the emergency communication procedures routes the communication to the most appropriate PSAP.

Some specific additional information is also expected to be provided in the emergency specification, as location information and language and modality preference indications, see ETSI TS 103 919 [6] for details.

An address suitable for callback is provided by the communication service to the PSAP in the initial communication.

ETSI TS 103 919 [6] explains how accessible emergency communications is performed based on the general packet based emergency communications interface specified in ETSI TS 103 479 [5].

5.3.2 Subscription of user equipment for use in a service

User equipment **shall** go through actions to prepare it for use with total conversation within a service. By that it gets provisioned for the service.

To work as a total conversation client, the user equipment **shall** be equipped with some software. The software may be preloaded before first delivery of the user equipment, or installed by the user as an app. Another way is to only preinstall Internet browser software, and then the communications program may be a web page with program contents loaded from a web server either when preparing the user equipment for use or when initiating communication with the user equipment.

The user equipment **shall** be subscribed to the total conversation service and receive a unique user identifier (see also clause 6.2.2 of the present document).

Registration in the service links a phone number or other user identifier to the network address of the user. That operation gives the service information on how to reach the user equipment with incoming communications to the user. If user preferences definition is supported, registration **should** ensure the user preferences are adhered to in establishing of the communication. For details see clause 6.2.4 of the present document.

Once the user equipment is registered with the service, it keeps the service updated by sending regular status updates every few minutes or when a change occurs. An example is a change of location affecting the routing of communication initiations.

NOTE: Additional pre-registration for getting access to specific services such as relay services may be beneficial to users or authorities. Authorities may want to limit resources spent on relay services, and therefore require pre-registration for relay service use. The use case is described in clause 4*.2.2* of the present document.

5.4 Communication protocols

5.4.1 General (informative)

The total conversation concept is defined on a high, communication protocol independent level. Specifications on how to implement total conversation in various communications technologies exist. IMS, SIP, and WebRTC [28] technologies offer, when the present document was authored, the state-of-the-art technical standards for total conversation, encompassing conversational video, real-time text, and voice.

IMS stands for the IP Multimedia System used by the mobile communications industry, standardized by 3GPP and profiled by GSMA.

SIP stands for the Session Initiation Protocol, specified by IETF RFC 3261 [17] and used in many cases for VoIP systems. IMS also makes use of SIP internally but according to its own profile of SIP.

WebRTC [28] is a technology that has the potential to run communications applications in Internet web pages, with support by web browser software, but also as installed software.

The following clauses specify the most important documents for implementing total conversation in these technical environments.

5.4.2 IMS MTSI

The IMS MTSI service makes use of the Session Initiation Protocol SIP for session establishment and maintenance as specified in ETSI TS 124 229 [7], also specifying media signalling in general by using the Session Description Protocol (SDP). When the present document was authored, MTSI is the predominant protocol used of multimedia communication in the IMS.

The details for all three media are specified in ETSI TS 126 114 [8], where the main presentation and transport protocols for real-time text are specified to use Recommendation ITU-T T.140 [16] for coding and presentation and IETF RFC 4103 [20] updated by IETF RFC 9071 [25] for transport. Mandatory video codec when video is supported is ITU-T H.264 [13], while H.265 [i.47] is in increasing use. For audio, support of AMR-WB, standardized as Recommendation ITU-T G.722.2 [12] is mandatory, while use of EVS standardized in ETSI TS 126 441 [i.26] is increasing. All these codecs can, with proper settings provide the quality required for total conversation.

5.4.3 IMS MTSI based on WebRTC and IMS data channel

A data channel for various data transmission tasks is included in the IMS concept. It is called the IMS data channel and is based on the WebRTC data channel technology. It is added to ETSI TS 126 114 [8]. That makes it possible to transmit RTT in that data channel and get error correction and other useful features directly from the implementation of the data channel in IMS devices. It is a desirable development to move to this form of communication for RTT, especially when video and voice moves to use the WebRTC transmissions for video and voice.

However, that requires conversion between two forms of RTT implementation in the IMS system. Standards are ready for that situation, so interoperability can be achieved between IMS devices implementing the traditional total conversation and the one with RTT in the IMS data channel.

Voice and video media are used as specified above in clause 5.4.2 of the present document.

5.4.4 SIP as used in VoIP

The Session Initiation Protocol SIP (IETF RFC 3261 [17]) has dominated real-time multimedia communication in IP networks for about two decades when the present document was authored. It is a protocol for session establishment while when the session is established, other protocols carry the media in the session. SIP is used in Voice over IP (VoIP) and in video telephony and video conferencing as well as total conversation.

The typical protocol collection for total conversation implemented in SIP is:

- Session control by SIP (IETF RFC 3261 [17]), secured by TLS (IETF RFC 8446 [23]).
- Media control by Session Description Protocol SDP (IETF RFC 8866 [24]), secured by TLS (IETF RFC 8446 [23]).
- Video e.g. by Recommendation ITU-T H.264 [13], transported by RTP (IETF RFC 3550 [18]), by packetization IETF RFC 6184 [21] and secured by SRTP (IETF RFC 3711 [19]).
- RTT by Recommendation ITU-T T.140 [16], transported by RTP (IETF RFC 3550 [18]), by packetization IETF RFC 4103 [20] updated by IETF RFC 9071 [25] for multiparty aspects and secured by SRTP (IETF RFC 3711 [19]).
- Voice by e.g. OPUS, IETF RFC 6184 [21], transported by RTP (IETF RFC 3550 [18]), and secured by SRTP (IETF RFC 3711 [19]). Another commonly available voice codec is AMR-WB also called Recommendation ITU-T G.722.2 [11] transported by RTP (IETF RFC 3550 [18]), and secured by SRTP (IETF RFC 3711 [19]). The voice codec Recommendation ITU-T G.722 [12] transported by RTP (IETF RFC 3550 [18]), and secured by SRTP (IETF RFC 3711 [19]) is suitable as a default to ease interoperability, has the required audio bandwidth, but requires 64 kbit/s transmission capacity, so it is better if any equally good codec can be selected which has lower transmission capacity requirements.

The Session Description protocol SDP is designed so that the parties involved in a session can declare multiple standards for each media, and the best common standard can be selected to be used in the session. This eases interoperability.

Passing routers and firewalls was originally problematic for SIP based communication but has been solved by a standard called ICE (IETF RFC 8445 [22]).

SIP based communication systems are based on SIP servers responding on requests to set up and perform communications sessions by SIP endpoints by users. The endpoints are usually smartphones, tablets or computers with the SIP endpoint software installed.

5.4.5 WebRTC

The last few years when the present document was written, web based communication has become standardized and starting to dominate the communication involving end user devices.

The communication between services is usually still made by SIP implementations.

The reason why WebRTC dominates in end user devices is that the main software needed for session establishment, and media coding and transmission is available in all modern web browsers, and therefore available in all modern end user equipment. Only relatively small programs handling the specifics of the communications session are needed to be fetched from web servers or installed in the end user device.

Communicating end user devices get by this approach the same communication program to execute in the session, which eases interoperability.

Communication in video and voice is very well supported in the browsers, using modern codecs and security. Also capturing and presentation of these media is well supported by the browsers.

Communication of RTT is decided to be done by using the WebRTC data channel. It has good support for reliability and security. It requires though that all input and presentation activities are done in application software or a web page program.

For communication in Web Technologies, W3C and IETF have created the WebRTC concept where the Reliable WebRTC Data channel concept IETF RFC 8831 [i.37] in its section 3.2 use case U-C 5 is recommended to be used for RTT. The use of WebRTC Data Channels in web pages and apps is specified in W3C WebRTC: Real-Time Communication in Browsers [28].

WebRTC changes the interoperability scene completely. The traditional mechanism for interoperability is that interoperating system components implement the same communication protocols in equal ways. With WebRTC, the communicating parties usually load the same web page containing the communication programs and let the web browsers execute the program for getting the communication going. Interoperability then depends on the Web browsers to act in the same way when they interpret the web page. Consequently, it is the web browsers that need to ensure interoperability when handling the web based communication.

It is therefore good for accessibility and for efficient implementation work if RTT is standardized in WebRTC and implemented according to the standard, but it is not critical for interoperability.

5.5 User equipment requirements

5.5.1 General requirements

User equipment intended for use for total conversation shall support the three-communication media inherent to any total conversation service, real-time text, video and voice. Software for handling the initiation of communication shall be operational when the communication is initiated. Software for handling the media flow shall be operational when the media flow is started. The software subscription routines are described in clause 5.3.2 of the present document.

The functional user interface and user interaction related requirements for total conversation clients are detailed in clause 6 of the present document.

In addition, any communication client **shall** also meet the applicable EN 301 549 [1] requirements:

- for ICT with bidirectional voice communication contained in EN 301 549 [1] clause 6; and
- for software, contained in EN 301 549 [1] clause 11; or
- for web, if the service is provided as part of the web page, EN 301 549 [1] clause 9.

In clause 5.5 of the present document, the hardware and software requirements for user equipment to be used for total conversation are specified, including the requirements to ensure compatibility with the relevant assistive technologies.

5.5.2 Compatibility of user equipment with assistive technologies

5.5.2.1 General

To provide a sufficient level of accessibility for all users, the user equipment used in total conversation **shall** ensure compatibility with assistive technologies. A catalogue of some relevant assistive technologies is provided in clause 4.2.4.

The compatibility with assistive technologies is managed at the level of user equipment (device, platform software or total conversation client software) and assistive technology rather than through any specialized protocols unique to total conversation communications services.

Assistive technology is an umbrella term encompassing any item, piece of equipment, service or product system including software that is used to increase, maintain, substitute or improve functional capabilities of persons with disabilities.

5.5.2.2 Accessibility services

Accessibility services are designed and provided to make the user equipment easier to use for people with disabilities. Examples of accessibility services are given in clause 4.2.4 and include screen readers or TTS/STT software.

The user equipment used for total conversation **shall** meet the requirements of clause 11.5.2 of EN 301 549 [1] to ensure compatibility with the relevant accessibility services.

Where the ICT supporting total conversation has closed functionality, that closed functionality **shall** meet the requirements in clauses 5.1 and 5.2 of EN 301 549 [1] to ensure access to the needed accessibility features.

NOTE: User equipment used for total conversation will not typically contain extensive closed functionality. Closed functionality refers to parts of systems or devices that do not allow users to employ or run third-party assistive technologies or otherwise customize the ICT beyond what is provided by the developer or manufacturer. With closed functionality, the accessibility services available to the open functionality of the ICT are not available, so the ICT needs to provide its own set of accessibility features to support user visual and audio interaction with the closed functionality of the ICT.

5.5.2.3 Local connection interfaces

User equipment used for total conversation **shall** provide at least one input and/or output connection that conforms to industry standards, allowing the connection of external assistive technology devices, as required by clause 8.1.2 of EN 301 549 [1]. These interfaces are crucial to allow a wide range of assistive technologies to be used seamlessly with total conversation services.

NOTE 1: The intent of this requirement is to ensure compatibility with assistive technologies by requiring provision of a standard connections in a non-proprietary format on ICT.

NOTE 2: The word connection applies to both physical and wireless connections.

NOTE 3: Current examples of industry standard non-proprietary formats particularly relevant in the context of AT are USB-C, Bluetooth® and Bluetooth® Low Energy Audio (LEA)/Auracast®.

5.5.2.4 Hearing aid compatibility

In addition to providing at least one standardized input/output connection, the user equipment **shall** be compatible with hearing aids such as telecoil and cochlear implants. Hearing Aid Compatibility (HAC) is defined by two separate ratings T-rating, indicating Telecoil coupling, and M-rating, indicating acoustic coupling.

NOTE 1: Four level scale is used for both T and M-rating:

- T-rating rating measures the inductive coupling capability of the user equipment when used with hearing aids equipped with a telecoil (T-coil), with T1 indicating the poorest level, and T4 indicating the best. User equipment rated T3 provide good inductive coupling, allowing clear sound transmission to the hearing aid's telecoil without interference or noise.

- M-rating measures acoustic coupling capability of the user equipment when used in microphone mode. It refers to the level of Radio-Frequency (RF) emissions from the user equipment and how likely it is to interfere with hearing aids in microphone mode. M1 rating indicates high levels of interference, and M4 indicates least interference and best compatibility with hearing aids. User equipment rated M3 are considered to have good acoustic coupling and meet the required limit for RF emissions.

According to EN 301 549 [1] clause 8.2.2, the user equipment used for communication **shall** provide at least T3-rating for the means of magnetic coupling to hearing technologies.

To ensure sufficiently low RF interference and sufficient quality of acoustic coupling, the user equipment **shall** meet the levels required for M3-rating.

The T/M ratings for the user equipment are determined by manufacturers who **shall** follow comprehensive testing and compliance processes based on the relevant standards. ETSI EN 301 489-1 [2] sets the general requirements for electromagnetic compatibility. Part 52 (ETSI EN 301 489-52 [3]) provides detailed requirements for cellular communication user equipment.

The magnetic field strength and inductive coupling quality, which directly relates to the T-rating of mobile phones, can be determined using the international standard IEC 60118-4 [9].

NOTE 2: To achieve the best compatibility between hearing aids and user equipment, it is also important that the hearing aids meet the minimum requirements regarding acoustic coupling. IEC 60118-13 [i.73] is a standard focused on the immunity of hearing aids to RF emissions, ensuring that hearing aids can tolerate certain levels of RF interference from mobile phones.

NOTE 3: While T/M ratings help in selecting the user equipment that is best for users with hearing aids, it is advisable that hearing aid users test the user equipment to ensure it provides adequate acoustic coupling for them.

5.5.3 User equipment requirements for provision of video

5.5.3.1 General

The generic user equipment requirements to support video communication for total conversation are contained in EN 301 549 [1], clause 6.5.

Any hardware deployed as a user equipment intended for total conversation shall have video input and output capabilities and include:

- Built-in camera or an external camera and an interface to connect a camera to allow for video input.
- Built-in display and/or an external display connected to a display interface to allow for video presentation.

The present document assumes that if nothing else is specified that the user equipment is intended to be useable for persons using sign language and when just one person is seen in the view.

An ideal view of a person using sign language capturing most signs is vertically from skull to stomach and horizontally slightly wider than seeing the shoulders.

The eye can resolve conveniently 60 points per degree of the view, as described in Visual Acuity [i.65].

Good sign language perception requires a resolution of about 320x240 pixels as shown in ITU-T H-series Supplement 1 [14].

The minimum width that a view of a signing person **should** have on a screen is therefore 320 pixels/60 pixels per degree = 5,3° horizontally. The used screen **should** therefore have at least 320 pixels over that width of view when viewed at a convenient viewing distance. This calls for different requirements per product category of products used as user equipment for total conversation.

It **should** be noted that many factors come into play contributing to a sufficient usability of video communication. Good lighting is one. A plain well contrasting background is another. The angle of view of the camera is another. The human capability to perceive sign language under different conditions also can be varied widely, so that language perception can be successful with much lower performance of the communications systems than shown below, but with increased stress and strain and risk for misunderstanding. Better performing systems or allowing the image of the signing person to take a larger part of the display than in the examples may also be appreciated by the participants and be experienced to be more convenient than following the minimal requirements.

5.5.3.2 Typical user equipment types

5.5.3.2.1 General on equipment types

For the study of suitable user equipment characteristics for sign language communication, user equipment is divided in a number of types with typical sizes and typical viewing distance. These types are smartphones, tablets, laptop computers and large video conference equipment.

5.5.3.2.2 Smartphone

A smartphone is typically a small device handheld or placed in a stand. It can be assumed to be used at a distance of about 350 mm from the face of the user.

The width of the area showing each signing person in a total conversation session **shall** then be 350 mm × tan(5,3°) = 32 mm. Over this width, the display **shall** be able to present at least 320 pixels. Display sizes and resolutions vary between smartphones, but an example can be that the display is 70 mm wide when held in the portrait position and **shall** then have at least 700 pixels over that width. When this is authored, most smartphones have at least 1 080 pixels on that width, so this requirement is feasible to meet.

The display area is needed for other purposes while the views of the communication participants are shown. There **shall** at least be room for an RTT presentation area and a self-view. An on-screen keyboard is also needed during times of communication via RTT during the total conversation session.

5.5.3.2.3 Tablet

Also tablets come in different sizes, and the reasoning here is for a typical device. A tablet used for total conversation is usually placed in a stand and viewed on a distance of 500 mm.

The width of the area showing each signing person in a total conversation session **shall** then be in the following width 500 mm × tan (5,3°) = 46 mm. Over this width, the display **shall** be able to present at least 320 pixels. Display sizes and resolutions vary between tablets, but an example can be that the display is 220 mm wide and **shall** then have at least 1 530 pixels over that width. When this is authored, most tablets have at least 1 940 pixels on that width, so this requirement is feasible to meet.

The display area is needed for other purposes while the views of the communication participants are shown. There **shall** at least be room for an RTT presentation area and a self-view. An on-screen keyboard is also needed during times of communication via RTT during the total conversation session.

5.5.3.2.4 Laptop computer

Laptop computers are also suitable as total conversation user equipment. A laptop used for total conversation is usually placed on a table with a built-in keyboard as stand and viewed on a distance of 600 mm.

The width of the area showing each signing person in a total conversation session **shall** then be in the following width 600 mm × tan (5,3°) = 56 mm. Over this width, the display **shall** be able to present at least 320 pixels. Display sizes and resolutions vary between laptops, but an example can be that the display is 400 mm wide and **shall** then have at least 2 300 pixels over that width. When this is authored, most laptops have between 1 940 and 3 400 pixels on that width. In this way, the requirement for sign language presentation in an area of 56 mm width may be slightly too high for some laptops and a need to let the sign language image of each participant take 65 mm for getting sufficient resolution for good perception.

The display area is needed for other purposes while the views of the communication participants are shown. There **shall** at least be room for an RTT presentation area and a self-view. The laptop display size is usually sufficient for either assigning an even larger part of the display for sign language participants, or for example for showing documents during the session.

5.5.3.2.5 Large video conference system

Large video conference systems are also suitable as total conversation user equipment when the sign language users take part in a larger meeting. A large video conference system used for total conversation usually has a large wall-mounted display with e.g. a width of 1 400 mm viewed on a distance of a few meters, say 3 meters as an example.

The width of the area showing each signing person in a total conversation session **shall** then be in the following width 3 000 mm × tan (5,3°) = 280 mm. Over this width, the display **shall** be able to present at least 320 pixels. With the example of a 1 400 mm wide display, the display **shall** have at least 1 400 mm/280 mm × 320 pixels = 1 600 pixels over its width. When this is authored, a common resolution on large video conference displays is 4k video, which has 3 840 pixels horizontally. Thus, the system in the example meets the requirements very well. In video conferences it may be common to show the image of more than one person in the picture. It should then be remembered that for good sign language perception, the view of each person **shall** be given at least 280 mm in this case, and therefore the system in the example could conveniently show 5 persons because 1 400 mm / 280 mm = 5 sign language users side by side.

The display area is needed for other purposes while the views of the communication participants are shown. There **shall** at least be room for an RTT presentation area and a self-view. The conference system display size is usually sufficient for either assigning an even larger part of the display for sign language participants, or for showing more than one participant, or for example for showing documents during the session.

5.5.4 User equipment requirements for provision of voice

The generic user equipment requirements to support voice communication are contained in EN 301 549 [1], clause 6.

Based on the performance requirements for voice perception as described in clause 5.2.3 of the present document, and considering the achievable performance of mobile devices, any hardware deployed as a user equipment intended for total conversation **shall** have voice input and output capabilities,

and **shall** include:

- Microphone and/or an interface to connect a microphone to allow for speech input in a frequency range of at least down to 250 Hz and at least up to 7 000 Hz.
- Transmission coding capability of audio at least down to 50 Hz and at least up to 7 000 Hz.

NOTE: This can be achieved with all audio codecs specified in clause 5.4 of the present document with proper settings.

- Speakers and/or an interface to connect a speaker to allow for speech output in a frequency range of at least down to 250 Hz and up to 7 000 Hz.

and **should** include:

- An interface to connect an external speaker or other listening device to allow for speech output in a frequency range of at least down to 100 Hz and up to 7 000 Hz.

Additionally, the user equipment **shall** provide a means to adjust the speech output volume level:

- over a range of at least 18 dB as specified in EN 301 549 [1], clause 8.2.1.1, and
- at least one intermediate step of 12 dB gain above the lowest volume setting as specified in EN 301 549 [1], clause 8.2.1.2.

5.5.5 User equipment requirements for provision of real-time text

Any hardware deployed as a user equipment intended for total conversation **shall** have text input and output capabilities and include:

- Keyboard or support an interface to connect to keyboard to provide a text input method.
- Dedicated display fields for presentation of the generated and received text or support an interface to connect an alternative display.

NOTE 1: In total conversation communication, RTT display fields are often presented simultaneously with video.

NOTE 2: According to the requirements of clause 6.2 of EN 301 549 [1] any user equipment used for voice communication that has text input and output capabilities, needs to support provision of real-time text to fulfil the accessibility requirements.

6 Functional requirements for total conversation clients

6.1 General

All total conversation clients are software and **shall** meet the accessibility requirements in clause 9 or 11 of EN 301 549 [1] depending on what type it is, and clause 5 of EN 301 549 [1] whenever applicable.

Clause 6 of the present document outlines additional functional requirements for total conversation clients.

6.2 Total conversation user profile

6.2.1 General

The personal profile of the user of total conversation **should** include:

- User identifier (a unique identifier assigned during subscription, see clause 5.3.2 of the present document, and used by other users for establishing connections with the user).
- User identification information (the information that the users want to be presented as their identification when they join the total conversation call).
- User communication modality preferences (the information about the users' preferred communication modalities) specified per communication direction.
- User communication language preferences per communication direction in combination with the modality preferences.
- Information on favourite/required relay service.

It may also include aspects such as:

- Preferred user interface layout including sizes, colours, etc.
- RTT presentation style.
- Alerting method.

6.2.2 User identifier

User identifier is the unique identifier assigned during subscription, see clause 5.3.2, and used by other users for establishing connections with the user. The user identifier may be numeric for number-based services and non-numeric for non-number based services. The user identifier **should** be shared with the receiving users during the communication establishment, but only when the communication is not anonymous.

6.2.3 User identification information

The total conversation client application **shall** allow the user to specify the information they want to share as part of their identification during the total conversation communication. Examples of such information include name, phone number or alias, any other additional information (such as for example affiliation).

Unless the communication is established in an anonymous mode, the user identification information **should** be shared during the communication session to identify each participant to the others.

The information **should** be modifiable for each communication session, and during any communication session.

6.2.4 User preferred communication modality

6.2.4.1 General

Total conversation provides an option to communicate using three different modalities: spoken, signed, and written, supported by three communication media: audio, video and RTT, respectively. By indicating the preferred or feasible communication modality, the users' communication can be facilitated accordingly.

NOTE: In practice, using total conversation the communication may involve other means than verbal, signed or textual communication. It may enhance the communication by allowing the lip-reading or use of pictograms. See also clause 4.1.2 of the present document.

6.2.4.2 User default modality of communication

Total conversation application client **should** provide an option for the user to specify their default preferred communication modality separately for input and output. Any communication session (incoming or outgoing) **should** be launched indicating these preferences. For more information about session initiation see clause 6.3 of the present document.

NOTE 1: The functionality is particularly important for deaf users whose native language is a sign language and for which video is indispensable, and the only viable alternative is RTT.

NOTE 2: The use of different modalities for input and output is common for people with hearing impairments, using speech for expression, but need RTT as complement or replacement for voice for perception.

6.2.4.3 User language preferences

The total conversation application client **should** provide an option for the user to specify their default language preference. The language capabilities and preferences are tightly coupled to the modalities. Indication of language preferences and capabilities for expression and perception in the different modalities form the information that can result in rapid establishment of functional communication.

6.2.4.4 Relay services needed for given communication modality

Total conversation application client **should** provide an option for the user to specify relay services that needs to be requested whenever communication is to involve a communication modality that is not accessible for the user.

NOTE 1: The functionality is needed to facilitate communication between the hearing and speaking users and the users with hearing impairments (deaf, hard of hearing, deafblind, or persons with speech disabilities). Relay services are needed to facilitate the communication, unless the parties find that they can use RTT as a common media for written communication.

NOTE 2: Additional recommendations on deployment of relay services in total conversation communications are specified in clause 6.4.3.2 of the present document.

NOTE 3: General accessibility requirements for relay services are provided in clause 13.1 of EN 301 549 [1]. ETSI ES 202 975 [4] provides detailed and comprehensive set of requirements for provision of relay services.

6.2.4.5 Preferred communication modalities for specific contacts

Total conversation application client **should** provide an option for the user to choose a preferred communication modality for selected contacts.

NOTE: For hearing sign language users, the video may be the preferred communication medium for selected contacts for use of sign language. This is why such selective setting of preferences is useful.

6.3 User control over the communication session initialization

6.3.1 Negotiating communication media to be used

All communication media **shall** be enabled for any total conversation communication. This will allow participants to freely use any combination of communication media for a particular total conversation communication session that they initiate or receive.

If participants' default or requested communication modalities are not compatible, clear information **should** be given to the parties that the communication cannot be established with the given preferences. A common media **should** anyway be identified and proposed to be used to establish the communication session, allowing the participants to identify a common communication modality.

6.3.2 Identity of the communication participants

All participants in a total conversation communication session **should** be associated with their respective distinct identification.

As required by clause 6.3 in EN 301 549 [1], the participant's identification (caller ID) **shall** be retrieved from the user's contact list and made available in text that is programmatically determinable, unless functionality is closed.

If the participant's identification is not available from the user's contact list, the identification **should** be generated automatically - using the participant's identification associated with their account (such as the name, phone number or alias, any other additional information, e.g. affiliation) and shared during the session as required in clause 6.2.3 of the present document.

6.4 Conversation facilitation

6.4.1 General

Multiple communication media in total conversation services allow for better facilitation of the conversation among participants who speak different languages or have other needs than the voice or video only bidirectional communication. The participants may adjust on their own the available communication media to facilitate the communication (clause 6.4.2 of the present document). Additionally, the communication may be further facilitated by assisting services, providing additional assistance to ease the communication (clause 6.4.3 of the present document). However, in the context of multiparty communications specific considerations are needed for facilitation of individual contributions (clause 6.4.4 of the present document).

6.4.2 Flexible choice and adjustment of media of communication

The total conversation client application **shall** provide the user with the functionality to freely switch on and off their use of real-time text, video and voice.

Additionally, the user **shall** be able to freely switch on and of the real-time text and video transmission.

Methods to switch on and off the various communication media **should** be implemented in an intuitive and accessible way. It is recommended to conduct user testing to ensure that the solutions meet the needs of all users, deploying the recommendations of EN 17161 [i.6].

NOTE: Many multimedia communication implementations have a background in voice communication and do not perform well in sessions without voice. Therefore, it is best practice to not omit voice communication technically from the transmission path, but just turn off audio input and output when the medium is not desired.

6.4.3 Assisting services for communication between users of incompatible communication modalities

6.4.3.1 General (informative)

Total conversation provides good opportunities for users to find modalities they can use for the communication, even when they prefer different modalities. However, there will be cases when it is not possible or not suitable for the users to adapt to common modalities.

In such cases communication assisting services are needed to allow for the communication between users who do not share any common communication modalities. Examples of such cases include:

- Communication between a hearing person (vocal language user) and a deaf person (sign language user) when they cannot use or feel uncomfortable using RTT.
- Communication between persons speaking different languages.

6.4.3.2 Relay services

Relay services are assisting services enabling users of different modalities of communication e.g. text, sign, speech, to interact remotely through ICT with two-way communication by providing conversion between the modalities of communication, by a human operator or by other means.

It is best practice to meet the applicable relay service requirements of ETSI ES 202 975 [4].

6.4.3.3 Translating services

Translating services are assisting services enabling users who speak different languages to communicate by providing a translation of the communication content into the language that is comprehended by each of the users. The services may involve a human translator, or be based on an automated translation technology, typically involving also TTS and STT conversions.

6.4.4 Facilitation of individual contributions in multiparty communication

Meetings with multiple participants **shall** provide a system for turn taking by requesting and giving "the floor", commonly indicated by a "hand raising" tool. This tool **shall** be used for contributions in all modalities.

For voice, it is a maximum of five participants is assumed for effective unmanaged conversation. In case of voice only communication involving two to five participants, the situations of input collisions are typically naturally avoided. However, for meetings involving more than five contributors, the turn taking mechanism **shall** be provided to mitigate the risk for too many speech collisions and hesitation because of that risk.

Effective turn taking management is critical for participants who contribute using real-time text.

Whenever there is at least one participant using real-time text for communication and at least one participant using voice or sign language for communication in a conversation with more than three participants, the communication in the meeting **shall** be managed to ensure that no contributions, regardless the media they use, are provided simultaneously.

It is challenging to process several communication inputs originating from multiple media occurring at the same time. In such situations, typically audio or video input easily dominate over the RTT input. If provision of input is not managed properly and RTT is provided simultaneously with communication input in other media, the task to consume all input in RTT and the other media becomes impossible for most users. For visually impaired participants who rely on text-to-speech conversion for accessing the written input, reading the RTT while listening to other parties in the communication may easily become an impossible task.

NOTE 1: One strategy to handle the RTT communication input on equal basis as input communicated in voice or video, is to have the RTT input read-aloud. For more details, see clause 6.5.3 of the present document.

NOTE 2: Written input in a form of complete text messages (placed in a Chat or similar channels) **should** be maintained available to allow participants to share input during the meeting that is not critical to be shared in real-time.

6.5 User interface considerations

6.5.1 General

The use of multiple communication modalities, and participation of users with different communication needs, requires specific considerations regarding the organization of the user interface elements in a total conversation client.

6.5.2 Presentation of contributions in total conversation communication

6.5.2.1 Active participant visibility and indication

In case of multiparty calls, the active participant who is communicating **shall** always be clearly indicated when active. In order to achieve that the following requirements specified in EN 301 549 [1] need to be met:

- Clause 6.2.2.2 concerning indication of participants who are actively communicating.
- Clauses 6.2.2.3 and 6.5.5 concerning indication of visual indication of incoming audio.
- Clause 6.5.6 concerning indication of total conversation participants who are active using sign language.

Ideally, the voice and sign language participants **should** be visible to increase eligibility of their communications, in particular to support lip-reading.

It is usually not essential that the RTT participants are visible for understanding of the text. But there may be other reasons for the RTT user to be visible, for example to convey body language or facial expressions.

6.5.2.2 Visibility of assisting service participant

Assisting service participant (i.e. a signing relay service interpreter or a language translator) **shall** by default be visible when conveying the interpretation or translation of the relevant input to the conversation participant that needs assistance.

NOTE: The participants who need assistance will change depending on who is taking active part. When signing participants are signing, the interpreter will be conveying their input to hearing participants in spoken language. But when speaking participants are talking, their input will be signed to sign language participants. Analogously, the participants needing assistance will also change in case of spoken language translations.

6.5.2.3 Presentation of real-time text

Real-time text **shall** be placed in a position near or below the main content of the screen. If feasible, RTT **should** be linked clearly to the texting participant, and without obstructing the view of the other visual information.

The RTT display shall meet the requirements of clause 6.2 in EN 301 549 [1]. In particular:

- The RTT presentation **shall** be grouped in readable sections of text collected from each participant sending text, as required by clause 6.2.2.1 in EN 301 549 [1].
- The presentation of text from the different parties **shall** not be merged on character level even if text is sent simultaneously from multiple participants. Different presentation views are possible. A common requirement on them is that the relative order in time when the text was received **shall** be approximately presented, as required by clause 6.2.2.4 in EN 301 549 [1].

- The content of the RTT communication shall be available for review during the communication, as required by clause 6.2.2.5 in EN 301 549 [1]. The review after the communication has ended would be possible only if the communication is recorded.

NOTE: Challenges in navigating user interfaces with assistive technologies (such as braille displays, screen magnifiers, or screen readers) are also relevant when interacting with RTT. To support effective navigation, it can be helpful to:

- provide clear notifications that new RTT has been received;
- make it easy for users to switch between RTT sources, including input from the local user.

Users relying on AT (e.g. deafblind users) may also benefit from accessing content from at least two sources in separate but adjacent areas of the display to facilitate easy reading of different RTT sources.

Further information on the presentation of real-time text is available in Recommendation ITU-T T.140 [16] Appendix I, ETSI TR 103 708 [i.13], clauses 7 and 8, and IETF RFC 9071 [25], section 4.1.

6.5.2.4 Visibility of captions or subtitles

Captions or subtitles **should** be placed in a visible position near or below the main content of the screen, if feasible, linked clearly to the active participant, and without obstructing the view of the other visual information.

NOTE: In case of automatic transcoding of speech or sign language to RTT, the transcoding process usually need to hear the full word before it can be written. This will result in a delayed presentation of the transcoded communication (see clause 5.2.4.3 of the present document).

6.5.3 Provision of automated modality conversions

At the time of drafting of the present document, automated modality conversion technologies are available between speech and text.

Methods for automated text-to-speech conversion are well developed and available now in multiple languages. They **should** be provided to total conversation users in order to facilitate communication between RTT and the participants that cannot read the RTT input. Still, they **should** be implemented in a way that does not to create any interference for screen reader users. Active use of automated text-to-speech functionality has been indicated as one possible strategy to facilitate RTT input in the context of multiparty communications (see clause 6.4.4). However, each user should be able to disable the read aloud for RTT locally, to avoid the interference with screen reader output, or to adjust the communication to their own needs and preferences.

Due to the very nature of RTT, the written text will not come in sentences but in small pieces as it is typed. Therefore, the read-aloud stream may be experienced as being too choppy and difficult to understand. The automated TTS tool **should** provide a setting for generating speech only for complete phrases.

Methods for automated speech-to-text conversation has reached a usability level that makes it being a rapid and easily applied accessibility feature during voice communication in many languages. The correctness varies depending on many factors. Speech-to-text functionality **should** be provided to total conversation users.

7 Security and privacy of total conversation service

7.1 General

Total conversation service as any electronic communication service raises both security and privacy concerns as the communication involves sensitive personal interactions and data. Those concerns need to be addressed for all total conversation communication media, audio, video and real-time text, and include:

- Total conversation service security concerns protecting the technical integrity of the service and preventing unauthorized access.
- Total conversation privacy concerns ensuring data confidentiality and user consent.

The ultimate security of total conversation communication will also depend on the overall security of the networks used for the communication.

Already existing systems upgraded to become total conversation systems **should** continue using their security measures and add the extra media securely. Clause 7 of the present document **shall** be seen as guidance for the requirements relating to implementing total conversation service.

7.2 Security requirements

7.2.1 General

It is convenient to consider the security requirements using the Confidentiality, Integrity and Availability (CIA) model indicating that ICT security needs to be maintained with respect to the three aspects. The basic requirements for total conversation service for each of these aspects are outlined in the subsequent clauses. Clause 7.2.5 of the present document contains technical standards relevant for implementation of the security requirements for total conversation service.

7.2.2 Confidentiality

7.2.2.1 General

Confidentiality is needed to ensure that whatever is sent over the communication network is only received at the endpoint and by the intended recipient.

Data encryption ensures that session information and voice, video and text transmitted during total conversation communication are only accessible by intended participants and are not intercepted by unauthorized, third parties.

Authentication confirms the identity of total conversation service users.

7.2.2.2 Data encryption

The fundamental strategy to achieve confidentiality of communication, is to use encryption. All transmitted and stored data **shall** be encrypted using robust encryption standards to prevent unauthorized access. This includes end-to-end encryption for all communication media; real-time text, video, and voice. Users **should** be assured that their private conversations remain confidential, with no risk of interception by unauthorized parties.

NOTE: Encryption is critical to address the eavesdropping concerns. These concerns are significant, especially for sensitive conversations such as communication that involves bank transaction.

7.2.2.3 Authentication

7.2.2.3.1 General

Authentication mechanisms **shall** verify the identities of users accessing the total conversation service. By using authentication, together with session and data encryption in clause 7.2.2.2 of the present document, users can be confident that their communications are protected from unauthorized access.

NOTE 1: Authentication helps to safeguard the users against cybercrime where attackers by impersonating legitimate users, gain access to sensitive data or take over their accounts, performing actions as they were the authorized user. Examples of cybercrimes resulting from unauthorized access include transfer of the users' financial resources to external accounts, financial frauds, corporate espionage, medical records theft, or hacking social media or email accounts causing a range of troublesome implications for the users (including reputational damage or financial losses).

NOTE 2: For cases where extra authentication is required for example in scenarios where there is widespread problem enabled by impersonation in telephone networks such as voicemail hacking or robocalling, the Secure Telephone Identity Revisited (STIR) standards are applicable as described in IETF STIR standards documents [i.43].

The user **shall** be provided with the functionality to adjust the desirable level of authentication for the general use of total conversation service. Most relevant authentication schemes include:

- Device Based Authentication (e.g. Trusted Devices): Users register trusted devices that are recognized during authentication, enabling a simpler and faster login experience on known devices. Especially useful for users who engage in total conversation across multiple devices - the trusted device authentication provides a seamless experience, reducing the need for repeated logins while ensuring security for frequently-used devices.

- Single Sign-On (SSO): The authentication process that allows users to authenticate only once and gain access to multiple applications or services without needing to log in separately. Useful in situations, where users want to use multiple services with a unified, simplified authentication process.

Multi-Factor Authentication (MFA) is a more advanced authentication procedure that requires users to present multiple forms of verification, such as something they know (a password), something they have (a mobile device), and something they are (a biometric), and may be warranted when using the service to perform especially sensitive communications such as communication with bank or medical services (though excluding emergency communications).

Authentication methods **shall** ensure that the user can use alternative means of identification, including non-biometric and biometric means as required by the relevant EN 301 549 [1], clauses 5.3, 9.3.3.8 and 11.3.3.8.

Authentication may rely on:

- What the user knows (knowledge) - the user knows and uses a password (openly or as a masked entry) or a memorized swipe path.

- What the user has (possession) - the user receives a dedicated verification code, or scans a QR code on an external device).

- What the user is (biometrics) - the user deploys a biological characteristic (such as face or fingerprint) to verify its identity.

The user device **shall** have a mechanism for authentication of any external devices to be connected and used in the communication such as headsets, similar to the requirement specified for home IoT in provision 5.1-5 in clause 5.1 of ETSI EN 303 645 [i.27].

NOTE 3: As an example: there **should** be limitation on the number of authentication attempts within a certain time interval when a user is trying to log on to user equipment and after a number of attempts the account is locked out.

7.2.2.3.2 Authentication during emergency communication

Total conversation service specified to be used for emergency communications **may**, depending on the local policy, allow such communication without any need for authentication.

NOTE 1: The user equipment documentation may specify this as needing to work with emergency communications by, for example: procurers, regulators, or product specifications.

NOTE 2: For mobile services, ETSI TS 123 167 [i.21], clause 7.4 specifies how this is accomplished.

7.2.2.4 Confidentiality of assisted communication

Assisting services are services invoked during a call, assisting the call participants with specific tasks in the call. Examples include language translation, relay service, or expert advice. The invocation and use of assisting services in total conversation related to emergency communications is specified in ETSI TS 101 470 [29] and ETSI TS 103 919 [6].

Such assisting services are also subject to confidentiality. The term communication assistance will be adopted according to ETSI ES 202 975 [4]. Communication assistant as a person working in a relay service with media conversion, as a human intermediary; including sign language interpreters for video relay services.

This communication assistant **shall** be subject to regulatory vetting sufficient for the service, as defined by the service. Communication assistance involved in total conversations need to be approved by authorities.

NOTE: Communication assistance for example is privileged to the user's bank details, and they can convey sensitive transactions on behalf of the user. During communication, the communication assistant maintains their independence

Communications assistants **shall** not disclose the content of any relayed communication, and they **shall** consider all transactions confidential, in compliance with the national law. The communications assistant **shall** not disclose what has been learned about the individuals and trade secrets as defined in clause 7.4 of ETSI ES 202 975 [4].

7.2.3 Integrity

Integrity mechanisms are needed to ensure that total conversation communication is accurate, reliable and trustworthy. Any data transmitted as part of the communication **shall** not be modified without detection and **shall** be received without any unauthorized alteration.

Data integrity mechanisms such as hashing and checksums, session tokens and Transport Layer Security (TLS) encryptions, as well as access control, are deployed as part of general communication protocols to ensure integrity of the communication in the total conversation service.

NOTE: Integrity mechanisms help to prevent threats such as malicious message tampering or Man-in-the-Middle (MITM) attacks where the attackers intercept the conversation (audio, video, or text) and either distort the information being exchanged or comprise their confidentiality.

7.2.4 Availability

Availability of a total conversation service corresponds to the degree to which the service is operational and functional for users when needed. Key aspects of availability for total conversation services from the human factors perspective include:

- Uptime/response time requirements: Depending on the nature of total conversation service, i.e. is it to be provided integrated in a widely available public service or a niche market service (for example, total conversation services for hard-of-hearing/deaf users), the uptime requirements might differ and are specified by different actors. The uptime is typically expressed as a percentage of the time the service is expected to be operational, excluding planned interruptions. For the users to recognize a communication service as available, realistic figures for uptime can be from 99,99 % or better for a widespread public service to 99,8 % or better for a smaller service. That means that total conversation service **should** ensure at least 99,8 % availability and not different than other services of similar size and type.

- User access: The service **shall** be available across various user equipment compatible with the given service (such as mobile phones, computers, or tablets) and the users **shall** be able to connect easily and consistently, regardless their location.

NOTE: For mobile technologies the viable user equipment needs to support UICC. For example, computers do not have typically UICC and may therefore not be able to support some of the service implementations.

- Data backup: Total conversation service providers **should** offer backup and storage of user data in order to ensure continued service availability in case of system failures that result in data losses. Users **should** give their explicit consent for storing their personal data. This is discussed in more detail in clause 7.3.1.2 of the present document.

Ensuring high availability form the user perspective requires a number of measures on the side of service providers that are outside the scope of the present document, but include among other provisions of mechanisms to address:

Redundancy and resilience: Service providers **should** ensure sufficient redundancy in infrastructure (e.g. multiple servers, data centres and network paths). Load balancing and resilient design of the systems allows to handle system performance during peak times and unexpected system failures.

Keeping total conversation service available if there is a loss of network is one of the measures that can be taken to increase resilience. This can be achieved for example by automatic switching to another network.

7.2.5 Standards to achieve security in total conversation

Table 7.1 provides an overview of relevant technical standards that **shall** be used to ensure confidentiality and integrity of total conversation service, depending on the communication technology deployed.

Table 7.1: An overview of relevant technical standards that shall be used to ensure confidentiality and integrity of total conversation service

Scope	Standard	Confidentiality	Integrity
Transport Layer Security Protocol	TLS (IETF RFC 8446 [23]).	x	x
Web real-time communications (Web RTC)	IETF RFC 9110 [26]	x	x
The Secure Real-time Transport Protocol (SRTP)	For audio and video, and for RTT when it is RTP based: SRTP (IETF RFC 3711 [19])	x	x
The Datagram Transport Layer Security (DTLS) Protocol	For audio, video and RTT: DTLS IETF RFC 9147 [27] Best practice is described in IETF RFC 8862 [i.66]	x	x
Session Initiation Protocol	IETF RFC 3261 [17] section 22 on authentication and 26 on security considerations, using TLS for transport security	x	x

7.3 Privacy requirements

7.3.1 Protection of sensitive personal data shared automatically

7.3.1.1 General

Sensitive personal data is data whose disclosure has a high potential to cause harm to the individual. Total conversation service like many other communication services collects personal data (such as passwords) and shares these data with the user equipment. Sensitive information about users that are stored on the user equipment and that are part of the user profile (see clauses 5.3.1.2.5 and 6.2 of the present document) are shared with total conversation service. All personal information is expected to be managed according to GDPR specification [i.74].

7.3.1.2 Storage of data

Manufacturers are expected to provide features within user equipment that support the protection of personal data stored on the device.

Sensitive security parameters such as encryption key for use by the user equipment, that are relevant for ensuring security of total conversation, **shall** be stored securely on the user equipment such as Universal Integrated Circuit Card (UICC) similar to the requirement specified for home IoT in clause 5.4 of ETSI EN 303 645 [i.27].

Manufacturers **shall** provide consumers with clear and transparent information about what personal data is processed, how it is being used, by whom, and for what purposes in the user equipment documentation.

NOTE: User preferences defined in clause 6.2 of the present document may be hard coded on the UICC and this need to be stored securely.

Data stored **should** be used only for the purpose for which it is defined.

7.3.1.3 Exchange of data

During initiation of service sensitive information are shared which include user identity, user preferences and communication participants (described in clause 6.3.1 of the present document). This information **shall** be transmitted securely, with the appropriate confidentiality, as described in clause 7.2.2 of the present document.

NOTE: Sensitive information transferred during service can include user preferences for service such as the capabilities of the most appropriate emergency service call-taker during an emergency communication (see clause 5.3.1.2.6 of the present document for details).

Where personal data is processed on the basis of user's consent, this consent **shall** be obtained in a valid way. Obtaining consent "in a valid way" normally involves giving consumers a free, obvious and explicit opt-in choice of whether their personal data can be used for a specified purpose. Users who give consent for the processing of their personal data shall have the capability to withdraw it at any time. This is also consistent with the legal requirements of GDPR [i.74].

If telemetry data is collected such as number of failed attempts logged into user equipment, the processing of personal data **should** be kept to the minimum necessary for the intended functionality and no more, according to the GDPR specification [i.74].

7.3.1.4 Deletion of data

The user **shall** be provided with functionality such that user data can be erased from the user equipment in a simple manner.

Mechanisms **should** be provided that allow the user to remain in control and remove personal data from services when the user leaves the service, user equipment and applications in some cases in the form of a delete function that removes the data from the immediate memory of the user equipment. When a user wishes to completely remove their personal data, they also expect retrospective deletion of backup copies.

The user **should** be provided with functionality on the device such that personal data can be removed from associated services in a simple manner similar to the requirement specified for home IoT in clause 5.11 of ETSI EN 303 645 [i.27]. Removing personal data "easily" means that minimal steps are required to complete that action that each involve minimal complexity.

Users **should** be given clear instructions on how to delete their personal data. Users **should** be provided with clear confirmation that personal data has been deleted from services, devices and applications.

7.3.2 Privacy of user communication

7.3.2.1 General

Privacy requirements with regard to sharing of sensitive information by the user during a communication will differ depending on communication media used, i.e. real-time text, video or audio. The requirements presented in this clause will be given with respect to the applicable communication input and output channels, namely:

- Visual output - used for input/output of RTT and output video.
- Visual input - used for input of video.
- Audio output - used for output of audio/voice.
- Audio input - used for input of audio/voice.

In addition, the privacy requirements for recorded communications, and assisted communication are included.

7.3.2.2 Privacy requirements for visual output

Visual output is essential for video as well as RTT communication.

In both cases, the users **should** be aware that others can see their displays - reading the text communication or seeing the video communication. These can be prevented by recommending the users to use screen shields that prevent sideways glances while providing a clear view for the intended users.

7.3.2.3 Privacy requirements for visual input

Visual input is relevant in case of video communication.

The users **should** be made aware that the video transmission is not selective, and their backgrounds are also captured by the camera and transmitted as part of the communication. The users' privacy may be maintained by providing the users with background images that cover the unintended background.

Users using sign language in total conversation communication will sometimes need to have the communication in closed rooms for privacy reasons because others in the same location can see what they say.

7.3.2.4 Privacy requirements for audio output

To ensure privacy of auditory output the users **shall** be advised to connect a mechanism for private listening such as earpiece or a personal headset as specified in clause 4.2.12 of EN 301 549 [1].

It is especially important to inform users of screen readers converting text based communication into auditory output about the privacy concerns and encourage use of mechanisms ensuring private listening to maintain the privacy of their communication.

7.3.2.5 Privacy requirements for audio input

Users **should** be educated to consider privacy of their communication when using auditory input. Especially users who are blind and visually impaired **should** be sensitized to the fact that their communication may be heard by third, unintended parties. To ensure and increase the privacy when sensitive personal information is communicated, the users **should** be encouraged to choose another communication media like text to convey the information.

7.3.2.6 Privacy requirements for recorded communications

Depending on the jurisdiction and applicable laws, an explicit consent may be needed from the communication participants whenever a communication is recorded.

Total conversation service **should** facilitate that by providing automated notifications to the call participants about the started recording.

NOTE 1: It is best practice to inform the call participants that the communication is being recorded, and about the purpose of the recording, and ask for their consent in a non-disturbing way.

NOTE 2: Educating users about the use of recording is essential to bring awareness and help to ensure that privacy of the service is maintained. In the service description special attention **should** be paid to the fact that images collected and stored are subject to GDPR especially if it contains third-party data [i.74]. This can be provided as part of the service documentation described in clause 8.4.3 of the present document, as well as in the form of alerts generated when the recording is activated.

7.3.2.7 Privacy requirements for assisted communication

Assisting services are services invoked during a call, assisting the call participants with specific tasks in the call. Examples include language translation, relay service, or expert advice.

Such assisting services are also subject to confidentiality. The term communication assistant will be adopted according to ETSI ES 202 975 [4]. Communication assistant is a person working in a relay service with media conversion, as a human intermediary, including sign language interpreters for video relay services.

Communication assisting services involved in total conversations need to be approved by authorities and are subject to regulatory vetting.

NOTE: A communication assistant for example is privileged to the user's bank details, and they can convey sensitive transactions on behalf of the user. During communication, the user maintains their independence - they **should** not feel confined by having the communication assistant involved as a third party in the sensitive communication.

Communication assistants **shall** not disclose the content of any relayed communication, and they **shall** consider all transactions confidential, in compliance with the vetting regulations and other relevant national laws. These requirements are contained in clause 7.4 of ETSI ES 202 975 [4].

8 Maintenance and Support

8.1 General

Maintenance is required to update the user's equipment and/or software to the latest version. A system's reliability depends on its maintenance and support.

8.2 Regular update

Maintenance involves keeping user equipment software updated on a regular basis. Developing and deploying security updates in a timely manner is one of the most important actions a service supplier can take to protect its users. It is good practice to keep all software updated and well-maintained.

An update **shall** be simple for the user to apply. An update that is simple to apply will be automatically initiated by the service similar to the requirement specified for home IoT in clause 5.3 of ETSI EN 303 645 [i.27].

Any automatic updates and/or update notifications **should** be enabled in the initialized state and configurable so that the user can enable, disable, or postpone the installation of security updates and/or update notifications similar to the requirement is specified for home IoT in provision 5.3.6 of ETSI EN 303 645 [i.27].

NOTE: It is important from a consumer rights and ownership perspective that the user is in control of whether or not they receive updates.

If an automatic update fails, then there **should** be a roll back option to the version before the update of the service app.

8.3 Integrated Diagnostic tools

The user equipment **should** contain integrated diagnostic tools that are easily accessible for maintenance and troubleshooting. The tools **should** help the users to diagnose problems impacting total conversation service with respect to network/service connectivity or important aspects of the user equipment (such as problems with input/output devices).

NOTE 1: Providing functionality that allows the users to test all communication aspects (including external devices) using the total conversation service with a test connection is advisable to help users to diagnose and resolve any problems in advance.

NOTE 2: In case of problems, automated and semi-automated repair mechanisms are encouraged.

The diagnostics tools **should** be customisable applications that make it easy for the user to detect and act on their output. The information about diagnosed problems **should** contain guidance on solving the issue, including among others links to the relevant resources such as service documentation or FAQ (see clause 8.4.3 of the present document), or shortcuts to helplines (see clause 8.4.2 of the present document).

8.4 Support

8.4.1 General

For total conversation service the relevant support services include but are not limited to: troubleshooting/support mechanisms, helplines offering customer and technical support, online training and FAQ resources.

8.4.2 Helplines

Helplines offering customer and technical support **should** facilitate accessible and efficient support for the service users.

All the interactions with service helplines **should** be designed to be accessible, allowing the users to freely choose the preferred modality of communication, and include relay services if needed, as specified in clause 12.2.3 of EN 301 549 [1].

If the request is connected to an AI assisted communication like chatbot, the user **should** be notified that they are communicating with a machine and not with a human.

If the call connects via an Interactive Voice Response (IVR) system, the IVR system **shall** meet all the relevant accessibility requirements from EN 301 549 [1], and in particular those specified in clauses 6.4 and 6.6 of that document. The user when connected to the IVR system **shall** be presented with the option of selecting a live agent as described in the book "Accessibility Playbook - Delivering accessible client service" [i.67]. It is advisable to have no more than four choice options per interaction level.

All live agents responding to the customer requests **should** be trained in handling the necessary accessibility needs such as being sensitive to the needs of persons of various abilities; ensuring clarity and enhancing understandability in the communication and allowing the customer, time to take in information, express themselves and make notes where necessary. Recommendations for agents are given for example in a customer service standard in Accessibility for Ontarians with Disabilities Act, 2005 [i.68].

In addition to information on service functioning, the support service **shall** also be able to provide information on the accessibility and compatibility features of the service, as specified in clause 12.2.2 of EN 301 549 [1].

8.4.3 Service documentation

Service documentation **shall** be made available in at least one of the following electronic formats as defined in clause 12.1.2 of EN 301 549 [1]:

- Web format that conforms to the relevant accessibility requirements of EN 301 549 [1], clause 9.
- Non-web format that conforms to the relevant accessibility requirements of EN 301 549 [1], clause 10.

Where documentation is incorporated into the user service interface, the documentation falls under the applicable accessibility requirements of EN 301 549 [1].

In addition to describing the general service functionality, the documentation **shall** list and explain how to use the accessibility and compatibility features as defined in clause 12.1.1 of EN 301 549 [1].

NOTE 1: Accessibility and compatibility features include accessibility features that are built-in and accessibility features that provide compatibility with assistive technology.

NOTE 2: It is best practice to use WebSchemas/Accessibility 2.0 [i.57] to provide meta data on the accessibility of the service.

Annex A (informative): Background study on the use of total conversation

A.1 Introduction

The purpose of the informative annex is to outline the background information on the use of total conversation. The materials gathered provide basis for identification of functional, service and accessibility requirements to achieve pan-European interoperable total conversation.

A.2 Use of total conversation

A.2.1 Introduction

Total conversation is a conversational communication concept containing video, audio and Real-Time Text or a subset thereof between two or more users of interpersonal communication services. Total conversation was first formally defined in year 2000 in Recommendation ITU-T F.703 [i.45] with the following definition:

- Total conversation service: An audiovisual conversation service providing bidirectional symmetric real-time transfer of motion video, text and voice between users in two or more locations.

The document goes on describing Total conversation and specifies two quality level profiles for it. Only the higher level is relevant today, called 4b-standard level, and specified as follows:

- Standard total conversation: with usable audio, good text and good motion optimized video.

The factor that differentiates total conversation from video telephony or video conferencing is the Real-Time Text (RTT) component. The immediate transmission of text while it is created enables the receiving parties to follow the thoughts of the sending party as they are expressed in writing. Real-time text (real-time text) was first called "text conversation" when its technical characteristics were first standardized 1998 in Recommendation ITU-T T.140 [16].

The remainder of this clause presents briefly use contexts for total conversation: use for interpersonal communication, with relay services, in emergency communication, and finally for multiparty, conferencing calls.

A.2.2 Use of total conversation for interpersonal communication

Total conversation significantly enhances the capacity of communication services. It has been fostered especially in the context of making the interpersonal telecommunication accessible for individuals who are deaf or hard of hearing or have speech impairments. Total conversation is essential for deaf persons for whom sign language is their mother tongue, video is therefore the only communication medium equivalent to voice for hearing persons.

Enhancing voice telephony with video has been contemplated almost since the advent of voice based telephony. There are several interesting descriptions of the early trials for videotelephony available online, see for example report by Ericsson [i.58]. Beyond the early 20th century futuristic trials, research and development efforts by various telecommunications companies in the US and Europe have all led to failures, because of inability to deliver the service with sufficient quality.

While technical capabilities prevented delivery of efficient video communications, other alternatives for voice telephony were explored, in particular - Real-Time Text communication. One of the oldest real-time communication tools is the OpenVMS PHONE Facility [i.55]. It allows real-time text communication in a split screen user interface among users on the local OpenVMS node or other nodes on the DECnet network. It was somewhat similar to some of the messaging facilities popular on the Internet today, but OpenVMS PHONE predates them and has remained virtually unchanged since the early 1980s and sends the text while it is created. OpenVMS PHONE supports sessions among more than two users but has been used in that manner relatively infrequently.

The software Talk is a Unix command-line chat program, originally allowing a form of Real-Time Text communication only between the users logged on to one multi-user computer, but later extended to allow chat to users on other systems [i.54]. Like OpenVMS PHONE, it uses a split-screen user interface and was popular in the 1980s and early 1990s.

Both tools however required PC based access and have not been widely adopted. With the uptake of mobile telephony, SMS or other forms of instant messaging emerged as alternatives for voice communication. However, while these communication means were more available for deaf and hard of hearing, they still provided substantially different quality of communication both with respect to the fact that for many deaf people text based language is a foreign language, but also with respect to the poor user experience of such message by message communication.

The study commissioned by Ofcom and performed by Plum Consulting in 2009 [i.49] compared the effective conversation speeds of various communication technologies. The conventional voice telephony was estimated to reach up to 170 words per minute (wpm) between hearing persons, and up to 150 wpm when the conversation was relayed using video relay between hearing person and sign language user. Alternative means of communication, including instant messaging, SMS or email, resulted in clearly inferior speeds of 30 wpm for messaging and 15 wpm for SMS/email.

Text based communication modes provide distinctively poorer communication speeds. They also do not guarantee synchronous, real-time conversation. The study did not consider direct video calls but given the video relay estimates adding a video communication improves significantly the effectiveness of communication for the deaf using sign language.

Further evidence for the importance of having video communication, to ensure equivalence between speaking and signing users, are provided by the estimates of general signing presented in clause 5.2.5 of ITU-T H series supplement 1 [14]. There, a sign language phrase containing 8 "words" is analysed. The phrase is estimated to be signed in 5 seconds, resulting in 100 words per minute. The estimate is lower than the one provided in the Ofcom study [i.49] discussed above, but the difference can be caused by the variation in speed of signing for different persons and situations. Also, speech speed is likely to differ and today the most common estimate for conversational voice speed is 120 wpm rather than 170 wpm.

While adding the video capability to the call is vital for the deaf, it is also important to acknowledge that it enriches the communication for all. With voice-only communication, nuances of non-verbal cues and body language are lost. Text-only communication further reduces communication capacities by significantly reducing the speed of communication, and by losing the auditory cues such that can be conveyed by the timbre or volume of voice. Total conversation, with its three basic modes of audio, video and real-time text, allows users to convey all these subtleties effectively. Facial expressions, gestures, body language, voice modulation as well as real-time text with emoticons - all provided as integral parts of the conversation add depth and authenticity to interpersonal communication. Users can seamlessly switch between real-time text, voice, and video modes based on their preferences or the nature of the conversation. This flexibility ensures that all individuals can communicate in a way that feels most comfortable and effective for them, promoting natural and authentic conversations.

Both video communication and text messaging are nowadays widely used. However, they have not been adopted as total conversation services, but rather as standalone services often provided as either pure video or pure text based services. Even though the real-time text technology has been available since 1980's, text based messaging services or chats provided as a part of video communicators are not implemented using real-time text.

Total conversation implementations have been largely limited to a specific context of applications for communication within the deaf and hard of hearing community. One of the first implementations of total conversation was done in 1997-1999 in Sweden in the "Framåt-2000" ("Towards-2000") project. Framåt-2000 was a project with participation of the Swedish Deaf Association and research and development participants and funded by the Swedish Ministry of Communication (KFB) and Allmänna Arvsfonden foundation. The project was established in response to the request from the deaf community. In the era of emerging video communication - the technology that addressed the core needs for sign language communication, deaf people found it as extremely useful to be able to have conversational sessions with a mix of the three media, combined as best suited for each communication situation. The real-time text function was also used for interoperability with telephone network based text telephones.

The total conversation technologies were subsequently advanced and used to establish first video relay services both in the USA and Europe. The next clause focuses on the use of total conversation in the context of relay services.

A.2.3 Use of total conversation together with relay services

Telephony services remained inaccessible for persons with hearing or speech impairments until introduction of telecommunication relay services in the 1970s in the US. The first telecommunication relay services were provided as text based relay services, where the deaf users could perform phone calls with the assistance of a communications assistant who relayed the conversation between a speaking party and the person with hearing/speech limitations, see Figure A.1 for illustration.

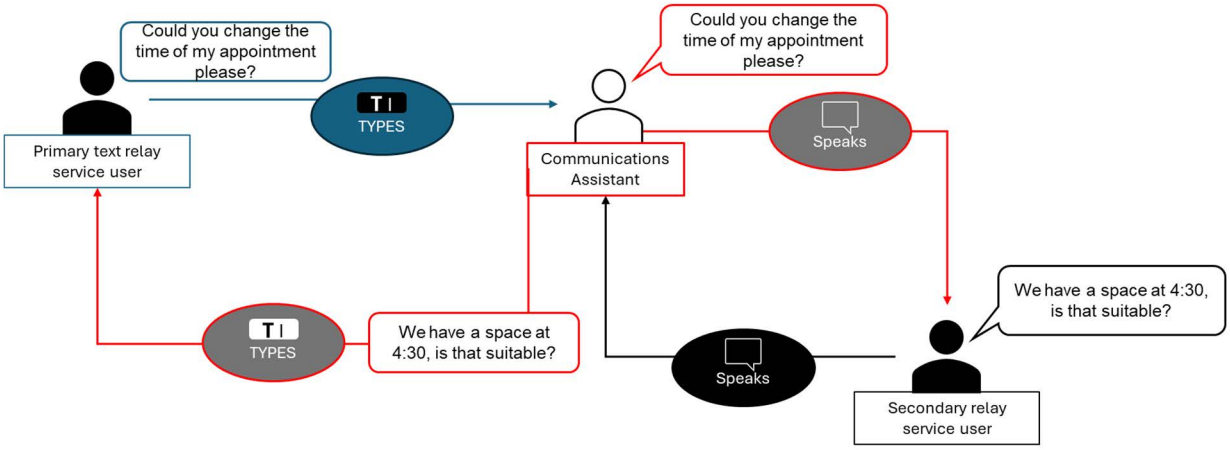

Figure A.1: Typical use of text relay services

In the 1970s through 1990s, the relay services were for use with text telephones, which were dedicated landline phones with a keyboard and small screen to type and read messages. Today, text based relay services are provided as IP services and enabled by a number of common communication devices such as smartphones, laptops, tablets, etc. Also, they are increasingly frequently provided as real-time text, or simply as a part of a total conversation service.

When total conversation is coupled with relay service access, it extends these capabilities to hard of hearing, deaf persons and persons with speech- or cognitive disabilities. Relay services act as intermediaries, facilitating communication between individuals who use sign language or text based communication and those who use voice. Deaf users, for example, can communicate using sign language, which is translated into voice for their hearing counterparts, and vice versa. This approach ensures that conversations are fluid and inclusive, breaking down the communication barriers that have traditionally separated the deaf and hard of hearing and deaf communities from the hearing world.

Figures collected during the research done for Ofcom by Plum Consulting in 2009 [i.49] confirm the benefits of having access to video relay.

It is estimated that the communication between a hearing person and a sign language user may be conducted with the speed of up to 150 wpm, comparing to only 30 wpm achieved with text relay. As noted earlier, sign language is for many of the deaf persons their mother tongue, and hence the video relay is preferrable as it allows them to communicate with their native language.

Given the speed of communication achieved with video relay this type of relay services provides most appropriate service for example in terms of the required by law equivalence in access to electronic communications. The video relay services increase the speed of relay service five-folds, bringing it up to the speed of regular voice conversation with 150 wpm. It is estimated that up to 170 wpm can be achieved in regular voice conversation, but an average, regular speeds is rather equal to 120 wpm. The speed of the text relay on the contrary has been estimated in [i.57] to yield only 30 wpm (increased up to 70 wpm for calls with voice carryover). With the improvement of texting technologies, in particular speech-to-text conversion technologies, the current text relay service often provide text relay achieving the speed of 50 wpm. Still, this is substantially lower than a regular conversation speed.

The analysis highlights also the benefits brought by captioned telephony. This service supports the conversation of hard of hearing persons who can speak but have no or limited hearing. Captions, texting what the hearing party is saying, help the conversation to be followed, at a rate almost equivalent to that of a voice call.

Both captions and availability of real-time text greatly enhance video communication. This is especially appreciated in the context of relay services, where some contents (like addresses or bank account numbers) are best suited for text based communication. That is why total conversation has been recognized as most appropriate for use in the context of relay services.

However, total conversation can bring substantial benefits to the communication at large, not only for the community of persons with disabilities. One such context in which total conversation is extremely useful is in the area of emergency communications. In everyday situations, the ability to freely switch between the different communication modes may be especially convenient and useful in noisy or quite environments.

A.2.4 Use of total conversation in emergency communication

According to ETSI TS 101 470 [29], the emergency service capable of handling total conversation emergency calls provides its users with a way to make emergency calls with total conversation and communicate simultaneously in a conversational way using available combinations of video, real-time text and audio. Callback from the emergency service with total conversation is also specified.

Extending modes of communication with emergency services allows for a quicker and more accurate communication. Video is likely to provide the emergency call takers with the background information allowing them to pose most appropriate further questions to gather all the necessary information. Real-time text can be indispensable in situations where the person in need cannot talk because of the experienced health condition, or possibly because of the external threat that is still present.

Total conversation clearly provides a higher quality communication platform for emergency communication in general. Still its implementations for emergency communications thus far have focused predominantly on its provision for deaf and hard of hearing persons.

While providing total conversation service for emergency communication can be technically feasible, it is important to note that establishing an efficient call may still be challenging. The emergency call takers can be expected to be able to handle text and voice in emergency communications, if it is in a language matching their competence. For calls with users in an emergency who prefer to use sign language, it is desirable but rarely feasible for emergency services to have call takers who are competent in both emergency call handling and in the national sign language. In such cases, it is important that the call is routed to the emergency service, and an assisting sign language interpreting service is invoked as a third party in the call to provide translation between sign language and voice.

A need to invoke suitable language assisting services as a third party in the emergency call also appears when the user in an emergency and the emergency call taker do not have any language competence in common.

In cases when there is a difference in language and modality preferences and capabilities between the user in an emergency and the available call taker, there is a need to assess the desired language and modality rapidly and invoke the correct language assistance in the call.

It may be very difficult and time consuming for a user in an emergency who needs to communicate in a specific sign language to explain that to a call taker only capable of handling voice and text in a language foreign to the user in an emergency. There are ways to signal language and modality preference in the call setup available for different technologies. Such signalling can be used for automatic or manual routing of the call to a suitable call taker or rapid invocation of proper language assistance. Even if such methods are used, the emergency services need to be prepared to sort out the language needs in manual ways, to handle cases when the call is made from a borrowed device, or the user has chosen to not indicate their language and modality preferences in settings for privacy reasons. A brief interaction in real-time text would in most cases be efficient for assessing the language assistance needs for cases when automatic means fail.

The language assistance may, when it is about translation between sign language and voice, be handled by the same organization as the relay service provided to the user for interpersonal communication, but there is a higher importance for rapid answering and 24-hour service provision than in regular interpersonal communication.

Another potential benefit of total conversation enabled emergency communication is to support the communication between different language speakers with automated translation of the real-time texting or even the automatically created transcript of what each party says. These benefits should be considered in the planning of future emergency services in Europe where citizens with different native languages travel across Europe and they can need access to emergency services in other Member States. In these contexts, robust access to emergency communication when roaming on another service needs to be ensured.

Feasibility of the total conversation concept in the context of emergency communications was demonstrated a decade ago in the REACH112 project [i.5]. REACH112 implemented a 12-month pilot in Sweden, UK, The Netherlands, France and Spain that allowed disabled users to communicate with each other, but also that allowed for a direct access to the emergency services. Using total conversation application emergency calls were established with the relay service interpreter that facilitated the communication. However, today there is still only a handful of total conversation implementations in Europe and elsewhere. These will be further reviewed and discussed in clause A.3 of the present document.

With the new European legislation, the European Accessibility Act, there will be requirements to provide total conversation emergency communications both for number based interpersonal communication services as well as the number independent communication services.

A.2.5 Use of total conversation for video conferencing

A.2.5.1 General

As indicated in clause A.2.1 of the present document, video conferencing can easily be confounded with total conversation as the video modality is also included in the total conversation. Still, the difference lies in the way text modality is provided and for the service to be total conversation, text needs to be provided in real-time, as it is typed.

Video conferencing has become a frequently used tool both at work and in everyday life for many users. Nowadays the video quality usually satisfies the needs of sign language users as well as users who need to combine hearing with support of lipreading for speech perception. The availability of real-time text together with video and audio in conferences enables rapid interaction in text without the waiting imposed by other forms of text communication. This can provide opportunities for persons who depend on text communication to participate in conferences on almost equal terms with others.

A.2.5.2 Real-time text in voice dominated conferences

In a voice-dominated video conference, availability of RTT can mean accessibility in nearly equal terms for persons who cannot participate fully by voice. An automatic or manually produced transcript of the speech can be provided in RTT (see clause B.1.8 of the present document for more information on speech-to-text technologies), and users who cannot use voice but can use text can communicate with RTT and have a chance to interact rapidly in the conference.

RTT can be a way to handle communication in person-to-person calls or purely text based calls. However, there is a risk that the RTT participants get ignored by the participants who communicate in voice or video. This risk is especially high with many participants in the conference. Also, presentation of documents in the meeting increases the risk that contributions in RTT gets ignored. Use of a hand-raising tool may be a way to get the floor in some meetings but not in others. Even if RTT would be noticed, there is a problem that creation of RTT by typing is much slower than speaking or signing, so a deaf participant would feel that the meeting is held waiting for the RTT text to be slowly completed.

On the other hand, RTT, and especially the functionality of automated transcription, possibly with automatic translation into the selected language, is extremely valuable especially in multiparty calls with participants speaking different native languages. Even if all share the knowledge of one language, captions can enhance the understanding of various contributions. Finally, RTT may also be useful to support participation in noisy environments that impede participants' ability to hear clearly or to contribute verbally.

A.2.5.3 Sign language interpreting in voice dominated conferences

Total conversation services also facilitate the inclusion of sign language interpreters, allowing deaf persons who prefer sign language to participate in a video conference in the way that allows them to both contribute to and receive the discussions most efficiently. In such cases, both the deaf persons and sign language interpreters participate in the conference and the interpreter signs what is said and speaks what the sign language users sign.

A.2.5.4 Accessibility needs in video conferencing

The arrangement for sign language interpreting in video conferences requires some features in order to provide proper accessibility. It should be possible for the sign language user to have a stable and sufficiently large image of the active interpreter. Sign language interpreters usually take turns in just a few minute intervals. The most favourable way of using interpreting in the conference would be to have the active interpreter always presented in the same place on the screen. This is different from a common way in video conferences to select only a few of the participants to be presented and move the images around according to who produces sound. When the correct approach of having a fixed placement for the interpreter is made possible, there is also a need for the arrangements to set up for use of sign language interpreters in the conference to be very simple.

A.2.6 Summary

The review provided in this clause has outlined how when video and real-time text communication technologies are provided together, this constitutes a total conversation service. The total conversion technology has evolved substantially over the years and is now capable of providing a mature and stable service.

While total conversation can commonly be confused with video communication, the service has much greater capabilities to support real-time communication providing not only video and audio communication modes, but also real-time text mode. As such, it provides the most comprehensive and universal communication service, with benefits for all the users in basic person-to-person communications, but also for emergency communications, multiparty calls, or calls requiring assisting services such as relay services.

Despite the clear potential of total conversation its application thus far has been limited to the use in specific context of provision of accessible communications services for deaf and hard of hearing. In the next clause, an overview is provided of the current use of total conversation in Europe and North America.

A.3 Current use of total conversation in Europe and North America

A.3.1 Introduction

According to the EAA [i.28], by June 28, 2025, all consumer electronic communication services that provide video in addition to voice communication and Real-Time Text (RTT) should be delivered as total conversation.

This clause reviews the actual use of total conversation in Europe and North America.

It is a common misconception to equate total conversation with a standard video communication service that includes a traditional chat feature. It is crucial to distinguish that while any video communication service integrated with RTT functionality constitutes a total conversation service, a video communication service offering a usual chat feature that does not operate as RTT does not qualify as total conversation.

Despite the widespread adoption of video communication as a ubiquitous tool in educational settings, numerous workplaces, and personal interactions, mainstream video services typically do not meet the criteria for total conversation, primarily because their text messaging capabilities do not align with RTT standards. To date, no widely-used video communication platforms that incorporate total conversation features were identified.

Total conversation continues to be a specialized service, predominantly utilized within the deaf and hard of hearing community for interpersonal communication. It also plays a significant role in relay services and emergency communication systems. This clause aims to give an overview of the deployment and utilization of total conversation across Europe and Northern America.

A.3.2 Current use in Europe

As of the time of the composition of the present document, the adoption of total conversation technology remains predominantly concentrated in a select number of European countries. This service is primarily accessible through installed applications, which are specifically deployed to enable video and text communication for individuals with disabilities. These services are number independent and are extensively utilized by video relay services.

The primary sources of information presented for the present document are the reports published by the European Emergency Number Association (EENA) [i.52], data gathered by the Body of European Regulators for Electronic Communications (BEREC) focusing on access to emergency services [i.3], and the expert knowledge of the project team members.

Table A.1 provides a detailed overview, distinguishing between peer-to-peer calls and video relay services. The 'Peer-to-Peer Calls' column refers to the use of total conversation or RTT applications that facilitate direct, interpersonal communication using these technologies, bypassing the need for a relay service operator. In contrast, the 'Video relay services' column indicates the use of relay services to facilitate both regular and emergency calls.

Table A.1 further to the countries where RTT is offered as an additional communication service. This often involves specialized emergency RTT applications. While most of these applications do not possess full total conversation capabilities, they do provide essential support for RTT.

Table A.1: Use of total conversation in Europe - status as of January 2024

Country	Function	Peer-to-peer calls	Relay services	Emergency communications via relay services
Sweden	**Total conversation**	**Yes**	**Yes**	**Yes**
	RTT	Yes	Yes	Yes
Norway	**Total conversation**	**Yes**	**Yes**	**Yes**
	RTT	Yes	Yes	Yes
Netherlands	**Total conversation**		**Yes**	
	RTT		Yes	
Denmark	**Total conversation**		**Yes**	
	RTT		Yes	
France	**Total conversation**	**Yes**	**Yes**	**Yes**
	RTT		Yes	Yes
Belgium	**Total conversation**		**Yes**	
	RTT		Yes	
Finland	**Total conversation**			
	RTT		Yes	
Germany	**Total conversation**			
	RTT		Yes	
Iceland	**Total conversation**			
	RTT	Yes		Yes
Bulgaria	**Total conversation**			
	RTT			Yes

The landscape of total conversation and RTT services is continually evolving, with ongoing efforts to enhance accessibility and inclusivity in telecommunications. This includes expanding the availability of these services beyond Europe and improving their integration with mainstream communication technologies.

Currently in Europe, total conversation is predominantly accessible through dedicated applications. The conducted review indicates that total conversation is available only in six European countries: Sweden, Norway, the Netherlands, Denmark, France and Belgium.

RTT is provided over relay services in Germany and in Finland. In Iceland, there is RTT available for peer-to-peer and emergency communications. And in Bulgaria, only emergency communications.

Interestingly, according to the recent BEREC report on measures for ensuring equivalence of access and choice for disabled end-users [i.3], nine out of 28 countries reported providing total conversation for relay services. These countries include Bosnia and Herzegovina, Belgium, Bulgaria, Cyprus, Czech Republic, Montenegro, the Netherlands, and Slovakia. The study conducted for the purpose of authoring the present document confirmed total conversation only in the Netherlands among the ones in that list.

Also, the information provided on BEREC's online catalogue of measures provided to facilitate access to emergency services for persons with disabilities [i.4] seems to provide not entirely accurate information. For example, for Norway only a text based relay access is reported, while the emergency services may be also accessed with the video relay service.

The observed discrepancy likely arises from the fact that the BEREC report collects information from National Regulatory Authorities for Electronic Communication specifically regarding services governed by electronic communication regulations. In some countries, for example in Norway, total conversation service is not provided as an electronic communication service but as a welfare service to facilitate access to interpreters for deaf and hard of hearing individuals. Only text relay service, which is a totally separate service, is provided as part of the electronic communication Hence, from the Norwegian NRA perspective relay services are provided only as text based services. Another reason for the discrepancies, is likely to be related to general poor understanding of what a total conversation service is, leading to inaccurate responses.

Dedicated applications that provide RTT/total conversation communication number independent services are provided to the users who need to access relay services. These users can then also use the apps for direct contact between each other. Still, the outreach is limited to the users that have the given application installed, allowing the communication within a given number independent service. As such the service cannot be seen as universally available.

According to the conducted investigations only in Sweden, total conversation services and apps are procured with a wording in the requirement specification saying that they "**shall** interoperate with other procured services". The result is that three service providers provide apps and services with interoperability between them and with the relay services.

Despite the development of services according to the same standards in different countries, there has only been occasional efforts within a project to ensure interoperability among total conversation services across these nations. In regions where total conversation apps are available, video relay services also relay emergency calls. Typically, deaf and hard of hearing are advised to use number 112, or a dedicated number or a button within the app to ensure their emergency call receives priority in the relay service. However, total conversation access to emergency services is not enabled for the general public [29].

A.3.3 Current use in North America

The conducted study identified the US as the only country where RTT is provided as regionally available communication service for all.

The Federal Communications Commission (FCC) in the USA has required both the wireless companies and manufacturers to make RTT available to all the users. The implementation has been gradual since on December 31, 2017 [i.29]. At the moment, all wireless networks support RTT in addition to voice for all calls. The RTT-functionality is also embedded as an integral part in the newly produced user equipment. handset that's RTT-capable. This is to the best of our knowledge the only implementation of the RTT service that makes the voice calls more accessible for deaf and hard of hearing.

It has also resulted in several user equipment available with the already implemented RTT. Using the feature from a regular mobile call app does not require any additional device, but the service needs to be supported by the carrier. Still, even in the USA where the service is enabled in the networks the service uptake has been rather limited and the RTT still remains a niche service used by the deaf and hard of hearing community.

As for total conversation, as in some of the European countries, the service is provided as an app based service to facilitate communication with relay services for the deaf and hard of hearing.

In Canada the situation is similar as in several European countries: Total conversation is provided as part of the video relay services, but RTT is only to be provided during 2024 for emergency communications.

A.3.4 Summary

The data presented in this clause indicate that total conversation service is offered only in five European countries (Sweden, Norway, the Netherlands, Danmark and France) and in the US. The clause reviews the use of total conversation in Europe and North America. In total it has been identified that 13 countries in Europe and the US and Canada provide total conversation service as part of their video relay service. Additionally, three countries in Europe (Iceland, Germany and the UK) provide RTT based relay services.

In all these instances, with a notable exception of the US for RTT, the service is not integrated into publicly accessible electronic communication platforms. Instead, it is delivered through dedicated applications designed specifically to cater to the communication needs of the deaf and hard of hearing community. These total conversation services are number independent and are utilized by users primarily for making calls that necessitate the support of relay services and for direct communication within the community. It is crucial to highlight that the availability of video relay services is not consistent, often lacking 24/7 accessibility. Consequently, these services do not provide a level of electronic communication equivalent to traditional voice telephony.

Only in the US, the RTT is provided as a publicly available service available for all electronic communications service users wherever they have access to voice telephony.

A.4 Performance requirements

A.4.1 Introduction

The quality levels of the different media components are defined in the Recommendation ITU-T F.700 [i.44] Framework Recommendation for Multimedia Services. The use in services, including total conversation services of these media and quality levels are specified in Recommendation ITU-T F.703 [i.45]. This framework not only defines the core principles of total conversation but also provides requirements for ensuring the delivery of real-time motion video, text and voice of satisfying quality during conversations. The key performance requirements are discussed below, drawing also on other sources.

A.4.2 Real-time text

The main performance requirement for real-time text is for the time it takes from the entry of a character or a small part of the text from a sending participant until it is visible to a receiving participant.

Clause A.3.2.1 of Recommendation ITU-T F.700 [i.44] on Audiovisual/Multimedia Services specifies "good text" in as T2:

- T2: Good text conversation quality characterized by:
  - Font support for all characters in ISO/IEC 10646 [i.64].
  - No more than 1 corrupted, dropped or marked missing character per 500.
  - Delay from character input in the transmitter to display in the receiver shorter than 1 s.

NOTE: Later requirement standards including EN 301 549 [1] limit the requirements on the font support to cover at least Latin-1, supported emojis, the replacement character and any subsets of characters for the languages in use in the intended implementation region.

Very brief research was made in 2005 as preparation for documentation in ETSI EG 202 320 [i.11]. It showed that in intensive text conversation, a delay between character entry and display of up to one second was found to not disturb the dialogue, while a delay of 3 seconds was slightly disturbing, and 10 seconds caused very disturbing problems in the interaction. It was also found that if transmission was made in bursts with over 300 ms of text entry each, then it was very important with a presentation smoothing the presentation of the characters in time. These findings confirm the requirement for less than one second delay for good text conversation quality requirement in Recommendation ITU-T F.700 [i.44], and the same in ETSI EG 202 320 [i.11].

The research participants also concluded that sentence-wise text chat is not suitable for intensive text conversation because of the disconnected feeling it causes for the participant awaiting next text entry.

It should be noted that the research was made with human input of text via a keyboard and the knowledge of that may influence the expectations of the receiving party. With the higher rate of RTT achieved with automatic speech recognition in conferences and the need to interact rapidly to have a say in a conference dominated by participants using voice, the level of acceptance for the longer delays may be lower.

The required performance in terms of the supported maximum rate of transmission of RTT characters with maintained delay requirements is 30 characters per second. This rate is intended to support all forms of human sources of text in a conversational setting. See ETSI EG 202 320 [i.11] clause 6.2.4. The performance for receiving RTT text should be at least 90 characters per second to support reception in multiparty sessions from up to three simultaneously transmitting participants. See IETF RFC 9071 [25].

A.4.3 Video

The performance requirements on video for conversational use over video communication is researched and documented in Recommendation ITU-T H-Series Supplement 1. "Application profile - Sign language and lip-reading real-time conversation using low bit-rate video communication", from 1998 [14]. For conversational use with one person in view, the spatial resolution required for good sign language perception is 352x288 pixels or more (320x240 is a common resolution that is also sufficient).

The required temporal resolution is to have at least 20 frames per second.

The end-to-end delay should not be more than 400 ms. Nothing is said about synchronicity with audio. Later research has shown that if video and audio are to be used together to support lip-reading there are requirements on the maximum asynchronism of these media. These requirements are specified in next clause.

A.4.4 Audio and audio plus video

Traditionally, telephone conversation has had a very limited frequency range between 300 Hz and 3 400 Hz. Sounds in voice communications have frequencies between 30 Hz and 18 000 Hz. Increasing the frequency range to between 250 Hz and 7 000 Hz improves clarity of speech very much. This is called Wide band audio. Hard of hearing persons benefit to a large degree to have this frequency range in the conversational communication. The frequency range between 50 Hz and 7 000 Hz is called wide band audio. The range between 50 Hz and 250 Hz does not add much to speech perception but has other values e.g. for person recognition.

The end-to-end delay of audio for conversational purposes should, according to the standards Recommendation ITU-T G.114 [10], be less than 150 ms for good conversational use and less than 400 ms to not cause problems with turn-taking. Recommendation ITU-T P.1305 [15], repeats these figures. For good lip-reading support of video with voice for persons with hearing impairments, video should not be more than 100 ms before or after audio. See Kozma-Spytek et. al [i.71] section 3.3.4.2 and Table 4.

A.4.5 Total conversation

The performance requirements on total conversation are the performance requirements of real-time text, video and audio expressed in the clauses above considering also the requirements on combined audio and video.

A.5 Critical aspects of total conversation provision

A.5.1 Introduction

As for any other communication services, also for total conversation its universal outreach is essential. One of the most universally available electronic communication services is a public number based telephony. The service has also been revolutionized with wireless mobile telephony taking over after the analogue landline systems. Ericsson® in [i.10] presents statistical data illustrating the decline in the use of fixed telephony and the uptake of the mobile telecommunications. For 10 years, starting from mid 1990s the number of subscriptions per 100 people reached 120 in Europe, 100 in the US and 110 worldwide. This contrasts with the fixed subscription which has declined in the same period by some 15 %.

The proliferation of mobile telephone subscriptions has been paralleled with the increased use of Internet over mobile networks. According to the data from OECD [i.53], the average usage of mobile data per user has increased five folds. The five OECD countries that have the highest monthly usage per user are all in Europe - Latvia (54 GB per month), Finland (46,7 GB), Lithuania (38,7 GB), Austria (36,6 GB), and Estonia (36 GB).

In 2022 the average monthly data used per mobile smartphone worldwide was estimated to 15 GB. Total mobile data traffic worldwide is expected to grow by a factor of around 3 over the next 5-6 years [i.9].

Mobile technologies become ubiquitous and provide a unique opportunity to enhance the traditional voice based telephony, providing alternative ways of communicating, including for example real-time text or allowing for the use of video communication services. The technologies to support and provide total conversation are here. Still, for the real uptake and adoption, an availability threshold needs to be reached. This is hindered by the lack of implementation of the common standards for total conversation service provision to allow the universal access to this type of communication.

Provision of universal access to the communications services requires interoperability. When communication services are interoperable, their users can establish communication sessions with each other and have information exchange in the common real-time media and modalities. This may also be achieved with services using web communication technologies when a service is made universally interoperable between the users in a way that only requires one party in a session to subscribe to the service.

The scope of interoperability of real-time communication have varied over time and varies today, especially when considering the three different modalities, i.e. voice, video and text.

From the 1960s to early 2000s, voice telephone users have had a fulfilled expectation that it is possible to reach any other voice telephony user in the world by means of a voice telephone and by using the international telephone number system for addressing the other users. Many technologies for voice telephony have been developed, and it has been common to provide them with interoperability between different types of technologies, so that voice calls could be made regardless of the technologies used in the end-user devices.

From early 2000s, various interpersonal communications services have emerged which have included real-time voice communication but did not provide universal interoperability with the global voice telephone network. Gradually, some of these systems allowed both call in and call out additions to its basic service to enable voice calls with users of the global number based voice telephone services, so the borderlines between the global voice telephone network and the new fragmented services is not entirely sharp. Most of these services have provided also text based communication channel, but these are typically not deployed as real-time text.

The new services typically use other addressing concepts than international telephone numbers. It is common with non-number based calling in the form of user name or an internet address in the form user@domain, see IETF RFC 3986 [i.34]. So, although some of the services facilitate interoperability with the global number based voice telephony, they do not provide any general interoperability with other, similar services.

Communication using other real-time media than voice has thus far not achieved the same level of interoperability between devices and providers as the number based voice telephony. Text based calls have been deployed using separate communication standards. Different standards for video telephony and video conferencing have been established for different network technologies with no interoperability between the networks and technologies. The user devices for video conferencing were predominantly delivered as dedicated units like the desktop video telephone or a larger videoconferencing system.

There were also some "soft phones" implemented as software applications, see [i.58]. Still, the communication was only possible between the units that worked over the same network type and adhered to the same standard.

The main technologies for the video telephony and conferencing systems were standardized by Recommendations ITU-T and included H.320 [i.75] for ISDN lines, H.324 [i.76] for analogue circuit switched lines, also used in 3G mobile communication, and H.323 [i.77] for packet switched communication.

Although H.324 enabled video calling in 3G mobile networks, the uptake of the service was low. Costs and culture were indicated as key factors putting people off from embracing video communication over the mobile phones [i.56]. Use by deaf persons for sign language communication had some penetration, but the achieved video quality was just around the threshold for providing some usability.

Once the generally programmable devices took over as the base for interactive communication devices, the situation changed. Now, smartphones, tablets and computers are commonly used for videoconferencing. They can perform as user devices and network components and switch between being used for different communication systems or even use different communication systems simultaneously. As such they enable video communication anytime, anywhere and with any user equipment/device. Still, one critical limitation is the persistent lack of interoperability between different video communication service providers.

From the discussion above a reasonably stable video communication technology at large has been available for the past two decades. Still, a wider adoption of this service has only started recently, supported by the increasing need for remote working and service provision. But it remains limited in its universality because of the lack of the interoperability between services. The remainder of this clause discusses in turn: the needs of the users with regard to interoperability of total conversation services, and next the mechanisms for interoperable text and video real-time communication services.

A.5.2 Interoperability of total conversation on human level

As indicated in the previous clause, a distinct feature of the currently available video communication services is their fragmentation and incompatibility, or lack of interoperability. The fact that video communication is restricted to its use only within a particular service, limits this service's universality. If users cannot call any other user in any communications service but require that the call receiving party is a subscriber of the same service, the video communication will lack the important outreach scope that is ensured with voice telephony.

While historically such fragmentation was explainable because of the technology barriers (different, incompatible networks), today those barriers are to a large degree removed as electronic communication networks are moved to IP based networks and implemented in digital devices which can switch between use of software for a multiple of different services. Still, fragmentation is maintained by lack of the interoperability between different service providers. The commercial operators of multimedia communication systems see little commercial benefit in making their services interoperate with those of their competitors. Instead, they try to add features and capabilities to their offerings and integrate them deeper into their wider hardware and software platforms with an aim of attracting users away from their competitors' platforms and tying them into their own. There are some initiatives at a European policy level to try to reduce the exclusionary nature of these isolated islands of communication capabilities.

Because of the lack of interoperability between commercially provided multimedia communication services, total conversation instead of becoming a mainstream communication service remains a niche service, used only for special purposes. This is troubling because universal adoption of total conversation would greatly contribute to increasing equivalence in access to communication services for all and especially for persons with disabilities. Indeed, the European standard on Accessibility requirements for ICT products and services, EN 301 549 [1], indicates in several clauses that to ensure accessibility wherever ICT provide means of voice communication, it should also provide an alternative communication mode.

From wider social and accessibility aspects, interoperable and universally available total conversation will give the opportunity to communicate in real-time even when the users have different capabilities or preferences in modality or language.

The basic total conversation service with its three real-time media enables a wider range of real-time communication modalities than voice communication. It makes it convenient for persons using sign language to communicate with other persons with sign language capabilities. It makes it convenient for persons who prefer text co communicate rapidly in text, also complemented with a view of the others in the session. And it can be used for voice communication. But in many cases it is not sufficient with these three real-time media to enable convenient communication. Modality conversion is needed to enable more use of the modality each participant in a communication prefers or masters.

Traditionally such modality conversions have taken place by manual action in relay services or interpreters invoked in the sessions. The ability of relay services to support multi-modal conversion is covered by EN 301 549 [1] as well as ETSI ES 202 975 [4] in more detail. Other standards on relay services are contained in clause B.1.6 of the present document.

Currently, when the present document is written, there is a trend to use automatic modality conversion as a complement to relay services for voice to text conversion. It may be fruitful to increase the efforts to make a wide range of modality conversion functions available for accessibility reasons. Such modality conversion services may also prove to be useful for other purposes, for example when users communicate using different languages. Automatically transcribed text (see clause B.1.8 of the present document for more information on speech-to-text technologies) can be automatically translated into different languages, facilitating the communication between users with different native languages.

Some of the currently available services/use cases where modality conversion is especially relevant include:

- **Text relay services** converting between real-time text and voice in number based voice telephony, usually converting in both directions but including variants for acting in only one direction.
- **Video relay services** converting between sign language and voice in number based voice telephony.

- **Captioned telephony relay service** adding real-time text to speech in one direction of a number based voice call.

- **Remote sign language interpreting in video and voice conferencing** where the meeting is using any of the common web based meeting services, with a possibility for the user to contribute by sign language translated to voice.

- **Remote text interpreting in video and voice conferencing** where the meeting is using any of the common web based meeting services, including a possibility for the user to contribute by voice or by text translated to voice.

- **Automatic speech to text function** (see clause B.1.8 of the present document for more information on speech-to-text technologies), adding text transcribed in real-time of what is said in a voice session or an electronic communications meeting. This functionality may work well for accessibility to voice based communication for persons who do not hear/understand well, but can speak.

NOTE 1: The potential group of users includes not only deaf and hard of hearing, but the same challenges are often faced by non-native speakers - captions are likely to improve greatly their understanding of the spoken language.

- **For persons who cannot hear and cannot speak clearly**, a text based communication is often the choice when they want to contact persons who cannot use sign language. Such communication today is often limited to an asynchronous chat service that do not establish real-time communication that would be equivalent to a voice call. real-time text addresses this deficit and allows for efficient text based calls.

NOTE 2: While a two-person voice/real-time text call or a purely text based call real-time text can be successful communication modalities, when the call involves more modalities, especially voice with multiple participants, there is a risk that real-time text gets ignored by the participants who communicate with voice. Also, presentation of documents in the meeting increases the risk that contributions in real-time text get ignored. Use of a hand-raising tool may be a way to get the floor in some meetings but not in others. Even if real-time text would be noticed, there is a problem that creation of real-time text by typing is much slower than speaking or signing, so a deaf participant would feel that the meeting is held waiting for the real-time text to be slowly completed. Other ways to create real-time text or speech rapidly is needed for this method to become accessible to an even higher degree.

Currently, the real-time text service is only available widely over mobile communication networks in the USA.

NOTE 3: The universally available real-time text is enabled by the requirement to adhere to a common real-time text implementation standard of IETF RFC 4103 [20], included together with voice and video in the mobile media standard ETSI TS 126 114 [8].

In Europe, real-time text is only available with dedicated apps (see Table A.1) used for relay service access and/or as a part of emergency communications in a few countries.

NOTE 4: For standards relevant to integration of total conversation with relay services see clause B.1.6 of the present document. For standards relevant to integration of total conversation with emergency communications see clause B.1.7 of the present document.

- **Automatic speech to sign and sign to speech functions.** There have been research and development going on for a long time in the areas of automatic creation and understanding of sign language. When it eventually reaches a maturity level suitable for use in sessions and meetings, it can be of benefit for deaf signing users to contribute rapidly to voice based meetings.

- **Interoperability with legacy telecommunications technologies.** Total conversation works well only in packet switched (IP) electronic communication environments. But the move to an all- IP electronic communications environment is not completed yet. If a session is established where total conversations users participate and voice telephone users in a circuit switched environment, automatic speech-to-text and text-to-speech technologies could be activated in the session either in the total conversation device, or in a network server, to provide some useful communication without engaging human operated relay services.

The fact that some of the early trials to introduce video communication to the marked has failed, and the rapid uptake of video communication in Internet communication services, leave the operators of number based interpersonal communication services sceptical to the potential of total conversation services. Still, from the perspective of values that are furthered in our societies, moving towards more inclusive communication environments seems inevitable. The user expectations on having access to ubiquitous multimodal communication service are still not so explicit. However, considering the forecasted changes in the way the communication services are to be used, with a notable and substantial increase of the use of video applications [i.9], it is not ungrounded to expect that over time users will require access to generally available communication that facilitates access to the three basic real-time modalities. It remains to be seen whether the universal telephony service will evolve to become a total conversation service, or possibly the dedicated proprietary services open for interoperability, including facilitation of communication with critical public services, such as emergency communications. In the next sections, some technical aspects are briefly reviewed that need to be addressed in order to achieve a wide interoperability of total conversation services.

A.5.3 Mechanisms for text based real-time communication services

A.5.3.1 Limited functionality in circuit switched analogue networks

Text telephony, the technology for a limited form of real-time text used in the analogue telephone network, was developed in at least six different variants, based on different modem technologies. Some of these were widespread within countries or small groups of countries, enabling good interoperability within these different country groups, but usually no interoperability between them. That caused significant inaccessibility for text telephony users when compared to the situation for voice telephony. An effort to create interoperability with a common modem type with backwards interoperability to the earlier ones was made, called Recommendation ITU-T V.18 [i.48], but it came too late to get any major positive influence on the situation.

A.5.3.2 Real-time text in packet switched networks

Full real-time text functionality was standardized on session and presentation level in Recommendation ITU-T T.140 [16]. It can be seen as the codec for real-time text. Codecs need agreed negotiation and transport mechanisms for each call control technology in which they are implemented. A currently dominating call control protocol is IETF Session Initiating Protocol SIP [i.18]. It is used in both Internet communication and in mobile communication. The currently by far dominating standard for initiating and transport of real-time text in the SIP call control environment is IETF RFC 4103 [20], which recently has got an addition to have a more precise specification on how to handle multiparty calling in IETF RFC 9071 [25].

Interoperability is always easiest to achieve if different devices and systems use the same standards. Interoperability in real-time text has been achieved in USA between users of different providers of mobile telephony using the 3GPP standards for SIP based calling in the mobile networks: ETSI TS 122 173 [i.18], ETSI TS 124 229 [7] and ETSI TS 126 114 [8]. These 3GPP standards for real-time text specify use of IETF RFC 4103 [20]. From January 2024 they also specify use of the update IETF RFC 9071 [25] for multiparty real-time text This update is when the present document was composed not included in the FCC requirements for implementation in mobile communications and not in the ATIS standards for real-time text in US mobile networks ATIS-0700029 real-time text Mobile Device Behaviour [i.1] and ATIS-0700030 real-time text End-to-End Service Description Specification [i.2], but the update is included in the standards for access to emergency communications in North America; NENA STA-010.3 [i.50].

Interoperability between providers of total conversation is achieved in Sweden between three providers of total conversation, by adhering to the standards referenced in public procurement specifications, namely [i.18], IETF RFC 4103 [20] and Recommendation ITU-T H.264 [13]. These providers use the addressing form used on the internet: user@domain. They use SIP for session control, H.264 for video, IETF RFC 4103 [20] for real-time text and several codecs for audio.

A group of European total conversation service providers provided international interoperability between users of total conversation of a number of different providers in a European project called REACH112 during 2009-2012. The addressing was based on international numbers, resolved in an ENUM [i.36] number data base to addresses of the form user@domain. The call control protocol was SIP and all used IETF RFC 4103 [20] for RTT and achieved interoperability.

A.5.3.3 Real-time text in web based communication

A rapidly developing technology for real-time communication causing a dramatically different way of handling interoperability has emerged. It is called WebRTC or communication in web technologies. In this form of communication, the task for handling interoperability is mainly placed on the web browsers in the end-user devices. In the basic form of WebRTC communication, users who intend to have communication fetch the same web pages with communication software to be run by the web browsers. The communication software utilizes communication functions in the web browsers. That may for example be audio and video codecs. The communication will likely pass through the same servers.

This means that the old sources of interoperability problems are gone, when servers having slightly different versions of software installed, or end user devices with different implementations of standards caused problems. Instead it is critical that the different web browsers handle the web page program and the communication streams in the same way so that the communication works. From this basic form, the mechanisms have developed so that the software can also be installed in the devices, but use the functionality of the web browser anyway, and detect in the server if there is any need to update the software to be sure that interoperability is maintained with the ones who run the web page based version.

In WebRTC [28], only audio and video are specified to use the real-time Protocol RTP for media transport. In regular SIP, RTT also makes use of RTP for transport. That caused the need of a new way to initiate and transport RTT in the WebRTC environment. RTT is specified to be using the WebRTC data channel when using the WebRTC technology. That is specified in the IETF RFC 8865 [i.40] standard for RTT in WebRTC.

NOTE: IETF RFC 8864 [i.39] was not working in the expected way at the time of writing the present document. Until this is resolved, IETF RFC 8865 [i.40] is proposed to be used in the way described in its section 1, by using the IETF RFC 8832 [i.38] procedures in the way specified in sections 3, 5 and 6 of IETF RFC 8865 [i.40].

Application in user devices use the W3C Application Program Interface (API) for WebRTC for performing the communication that can be provided through a web page as described in Annex B of [28].

A.5.3.4 Mechanisms for interoperability within technologies

The digital communication protocols used for real-time conversational use, are usually structured in addressing mechanisms, session control procedures, media negotiation and media coding and transport procedures. As discussed in the introductory clause, initially video communication was possible only within same technologies and equipment adhering to the same standards and protocols such as Recommendations ITU-T H.320 [i.75] or H.323 [i.77].

Today there are two predominant technologies: the Session Initiation Protocol (SIP) and WebRTC. For each technology, the basic procedures for addressing, session control, media negotiation and coding and transport procedures are specified so that they can be used in intended network environments. One example of network environment is the Internet including routers and firewalls with features within a common range. The procedures may also be complemented by procedures specifically intended for helping session establishment and media to traverse well through such routers and firewalls. For other networks, such as packet based mobile networks, there are other network aspects which need to be considered when applying the session control and media communication procedures.

The addressing mechanisms are usually based on the international telephone number plan, or an address based on only user name within a service, or a combination of user name and internet host address on the Internet for sessions both within a service and between services. Number based addresses are converted to an address in packet networks before use by the session control procedures.

The session control procedures find the way between the parties to be involved in the session, and let them negotiate which common media, codec and media transport procedures which are common to the parties in the session and therefore selected to be used in the session. Usually, communication devices have a multiple of codecs for audio and video, and the negotiation procedures are intended for selection of the best common codec. In this way it is possible to develop and improve communication gradually within a service or within services with interoperability.

The SIP technology is used for many communications services in the Internet and for packet based mobile communications services. For RTT, there are very few options. There is one codec specified to be used in all regular SIP based services, including packet based mobile services. That is Recommendation ITU-T T.140 [16] on the coding and presentation level, and IETF RFC 4103 [20] with addition of IETF RFC 9071 [25] for packetization and transmission.

In case of the WebRTC technology, it is possible to use SIP and the same codec negotiation procedures as for "regular" SIP but transported with other transport protocols. And RTT is specified to be transported in the WebRTC data channel using IETF RFC 8865 [i.40], which is different from what is used in "regular" SIP.

NOTE: IETF RFC 8864 [i.39] was not working in the expected way at the time of writing the present document. Until this is resolved, IETF RFC 8865 [i.40] is proposed to be used in the way described in its section 1, by using the IETF RFC 8832 [i.38] procedures in the way specified in sections 3, 5 and 6 of IETF RFC 8865 [i.40].

A.5.4 Mechanisms for interoperability between technologies

When a session is wanted between user devices using different session control technologies, then the most used method is to let the session pass a gateway for protocol conversion. The gateway is a server implementing both session control technologies and translating the characteristics of the session between the two session control technologies. Such gateways are commonly implemented without further standardization, while it is common to have sections about specific gateway considerations in standards for various aspects of sessions or media. Examples of standards which have gateway considerations sections for specific aspects are IETF RFC 8865 [i.40] for gateways between SIP and WebRTC technologies for RTT, IETF RFC 9071 [25] for multiparty aspects of gateways between RTT in SIP and WebRTC.

NOTE: IETF RFC 8864 [i.39] was not working in the expected way at the time of writing the present document. Until this is resolved, IETF RFC 8865 [i.40] is proposed to be used in the way described in its section 1, by using the IETF RFC 8832 [i.38] procedures in the way specified in sections 3, 5 and 6 of IETF RFC 8865 [i.40].

In some cases, the coding, packetization and transport of media can be carried through between the technologies with no or minor modification. This is preferable, because any changes in media coding, packetization and transport cause delays and introduce a potential for loss of media quality. However, in some cases to achieve interoperability changes of media coding, packetization and transport are inevitable and are done in a media gateway, often separate from the session control protocol gateway.

This way of efficient gateway handling can sometimes be achieved for interoperability of video and audio between the VoIP SIP technology and the WebRTC technology, when the same codec standards for video or audio are supported on both sides, because these technologies use the Real-Time Protocol (RTP) for both technologies. For RTT the text requires packetizing in new packets between these technologies, because RTP use for RTT is not available in WebRTC, and therefore the T.140-coded text needs to be repacked between RTP and the WebRTC data channel, As specified in IETF RFC 4103 [20] updated by IETF RFC 9071 [25] for RTP, versus IETF RFC 8865 [i.40] for WebRTC. A factor making the interoperability easy is however that both technologies use the coding and presentation standard Recommendation ITU-T T.140 [16], so on that level there is no need for transcoding.

A.6 Conclusions and future visions of total conversation use

The video communication and text chatting services have found widespread use today, especially after the pandemic period where most activities and communication was forced to be moved online. The services remain however fragmented and there is no uniform way to provide three-modal communication in a universal manner, like in the case of the voice telephony.

Among persons using of sign language communication, video communication has been the first choice as this is the only communication medium that allows these users to communicate in their native language. Both in the Northern America and in Europe services are provided to support video communication for sign language users, including provision of relay services. In addition, text relay is provided and these services in several cases are provided as total Conversation services with RTT integrated. However, the main deficiency is the lack of interoperability between the different total conversation services, inhibiting greatly their potential, but also limiting access of persons with disabilities to electronic communication services.

The review presented in the present Informative Annex A indicates that technology is there to support universal implementation of total conversation. Still, the technologies are clearly available today to deliver the services on a universal, global basis. The usage of data mobile technologies, that are essential for total conversation, has been growing exponentially over the last decade. According to data presented in Wytec International Executive Summary 2023 report [i.59] global data usage amounts at the moment to almost 110 000 PB per month, while in 2010 it has been estimated to be at the level of 279 PB per month. At the same time voice usage has increased from 147 to only 289 PB per month in 2023 since 2010. This means that the global mobile data usage exceeds the use of voice by an amazing factor of around 375.

The wide adoption of mobile data indicates that the potential for widespread adoption of total conversation is there. More importantly, both the historical data and the future forecasts indicate that video traffic has been and is likely to continue to be fastest growing application category when the overall mobile traffic is analysed, see for example data on mobile traffic for the period of 2012-2029 done by Ericsson® in their report [i.10]. Among the main drivers for video traffic growth the use of video for watching online content, sharing services, streaming are named. As the networks continue to adjust to the changing user behaviours, total conversation is likely to be even more easy to deploy. Given the expected extraordinary increase in the use of video, it is not ungrounded to expect that over time users will require access to generally available communication that facilitates access to the three basic real-time modalities. Video based communication has already been found to be yield more effective and productive work environments (see [i.60] for a review of the relevant survey results), albeit, as indicated in clause A.2.5 of the present document, the current mainstream video conferencing does not support total conversation. However, given the expected continued growth of video communications in the workplaces, the regulated introduction of RTT in the general public communications networks, already done in the US [i.29] and expected to be introduced soon in Europe [i.28], it is not unlikely that the users will require access to generally available communication that facilitates access not only to voice and real-time communications, but allows for adding the third communication medium - of video. Future trends in new telecommunication technologies and user behaviours and needs may also elicit the development of the various proprietary video communication services so that they support interoperability between each other, as well as access to critical public services, such as emergency communications.

The analysis in this Annex indicates that one of the barriers is the existing limited outreach of total conversation applications, largely limited to a dedicated use for communication purposes of the deaf and hard of hearing community. As emphasized throughout the present document, the differentiating factor for most popular video communication services from total conversation is the fact that they do not provide real-time text in addition to video and voice. With the implementation of the EAA, all the electronic communications services that provide video communication should be delivered as total conversation to conform with the EAA requirements. This may encourage video service providers to improve their services, making them even more accessible by providing them in total conversation format.

The other barrier that can be seen to limit the widespread adoption of total conversation is the profound fragmentation of the current video communication services, and their imminent lack of interoperability between the different video services. Clause A.5 of the present document reviews the mechanisms needed for total conversation provision. Clauses A.4.3, A.5.3.4 and A.5.4 of the present document show that there are established standards to provide successfully interoperable total conversation over modern data communication networks. Still, the implementation of total conversation as a universally available communication service is missing. The present document intends to contribute to solving this problem. It delivers a comprehensive guide on how to implement interoperable total conversation service, that will be of value for implementers and service providers of total conversation.

The study presented in this annex indicates that total conversation has several features that seem to be attractive enough for the service to become widely adopted and used, not only by users for whom it brings benefits of ability to communicate in native sign language. These include:

- Faster communication in critical, emergency situations when adding video communication supplements the voice communication, and the additional ability to use RTT can help precision in message formulation.
- Faster communication with RTT direct call, when voice communication is not possible to be established.
- Easier communication between persons speaking foreign languages - the functionalities related to automated captioning and/or translation are likely to make the communication between different language speaking users easier and/or more comfortable.

However, there is also a room for improvement of the current total conversation services, especially with respect to their RTT functionality. There are many good features in modern text chatting that would be valuable to have also in the real-time text communication. Some of these are:

- A brief notification at end of sentence.

- The sense of an eternal session without specific calling action.
- Start the voice or video session as an addition to the ongoing RTT communication. A habit is spreading that it is polite to ask first by text if it is suitable to have video or voice added.

NOTE: Notably two earlier ETSI standards - ETSI ES 201 275 [i.61] and ETSI EG 202 116 [i.62] - describe, as a privacy feature, how a user can configure their videotelephone to always answer a call in audio only, with the other end being notified that the receiving party's user equipment has video capability.

- It has become popular to have automatic speech to text included in video calls. It makes use of a real-time text mechanism. It would be useful to have that function totally integrated with the real-time text conversation so that some parties in a call can talk and some text.
- Include more rapid ways to produce real-time text:
  - Automatic sign to text.
  - Maybe someday automatic thought-to-text but with a rapid send/delete/edit action.
- Very setup of good support from sign language interpreting and captions in conferencing, possibly through the same connection as the sign language user. Today it can be very complicated to set up and maintain good support, where interpreters stay visible, etc.
- Become the presenter in a conference by starting to sign in the same way as voice users become presenters by starting to talk.
- Providing language and modality preferences automatically so that proper call agent can be reached, or proper interpreting can be arranged.

The material presented in this informative annex indicates that existing standards allow for implementation of universally interoperable total conversation, see clauses A.4.3, A.5.3.4 and A.5.4 of the present document. Also, technology proliferation is sufficient to support provision of such services. Moreover, the forecasts as presented by the study from Ericsson® [i.10] indicate that the user behaviour is likely to evolve in the way that it will be natural to expect easy access to video enhanced communications, the forecast indicated that the cumulative increase in video service in 2029 will be about 300 times more than it was in 2020 [i.9]. The main obstacle in enabling universal total conversation communication services seems to be in a imminent fragmentation of the services, their lack of interoperability. One barrier is likely be the fact that total conversation specifications and standards are currently scattered across multiple sources. Provision of one integrated standard that in a comprehensive manner outlines how to develop and implement interoperable total conversation service is clearly missing. It is this gap the present document is aimed to address.

Still, the different video communications service providers seem uninterested in achieving interoperability. A promising development is then the move to Web communication services by using the WebRTC technology, described in clause A.5.4 of the present document. Using this technology, interoperability can be reduced to that the users need to have a reasonably up-to-date web browser in their device. The caller needs to have a way to send a link to the callees, that appears as an incoming call. Answering the call makes the answering part fetch the same program for communication which is used by the caller and no interoperability problems should appear. The task is then reduced to provision of this way of calling in a secure user friendly and universal way. The number based addressing in the mobile communications systems can play an important part for this universal way of providing a link for a web based total conversation call.

With this way, the accessibility issues instead move to the problem that each call can provide a different user interface as it is provided by the web page linked to in the call. That will be a challenge for all communications providers to provide accessible user interfaces. It will also be a challenge for users with disabilities to rapidly orient in a possibly new user interface for each incoming call.

Annex B (informative): Related standards

B.1 Catalogue of standards relevant for total conversation services

B.1.1 Introduction

The present annex provides an overview of the currently existing standards regarding different aspects of total conversation provision:

1) Standards regarding user interface and functionality.
2) Standards for use in SIP for general IP networks and mobile communication.
3) Standards for use in Mobile Multimedia Telephony from 3GPP, GSMA, and ATIS.
4) Standards for Web Technologies.
5) Standards for use with relay services.
6) Standards for emergency communications.

B.1.2 Standards regarding user interface and functionality

Standards number	Title	Summary
EN 301 549 [1]	Accessibility requirements for ICT products and services	The document specifies the functional accessibility requirements applicable to ICT products and services, together with a description of the test procedures and evaluation methodology for each accessibility requirement in a form suitable for use in public procurement within Europe. The document is intended to be used with web based technologies, non-web technologies and hybrids that use both. It covers both software and hardware as well as services. It is intended for use by both providers and procurers, but it is expected that it will also be of use to many others as well. It has a clause which contains detailed accessibility requirements on real-time text. It also has requirements on total conversation, relay services and emergency communications related to RTT and total conversation.
Recommendation ITU-T T.140 [16]	Protocol for multimedia application text conversation	This Recommendation specifies a text conversation protocol. The intention of this protocol is to be a common presentation level suitable for straightforward real-time text conversation in multimedia services and in text telephony. It is based on ISO/IEC 10646 [i.64] Universal Character Set 16-bit characters and features character-by-character transmission and a limited set of presentation controls. It is meant to be easily applied wherever there is a data channel available to carry the protocol.
Recommendation ITU-T T.140 Addendum 1 [16]	Protocol for multimedia application text conversation Addendum 1	This addendum adds a marker for missing text to Recommendation ITU-T T.140 [16].
ISO 9241-20:2021 [i.7]	Ergonomics of human-system interaction - Part 20: Accessibility guidelines for Information/Communication Technology (ICT) equipment and services.	This part provides guidelines for making ICT equipment and services accessible, including those used for communication between people, ensuring that users with a wide range of abilities can effectively interact with systems.

Standards number	Title	Summary
ISO 9241-110:2020 [i.8]	Ergonomics of human-system interaction - Part 110: Interaction principles.	This part provides principles for the design of interactive systems that can be applied to systems facilitating communication between individuals. It focuses on general principles such as suitability for the task, self-descriptiveness, controllability, and conformity with user expectations, which are critical when designing interfaces for communication tools.
EN 17161:2019 [i.6]	Design for All - Accessibility	This standard addresses the concept of Design for All in the context of developing accessible and inclusive products, goods, and services.

B.1.3 Standards for use in SIP for general IP networks and mobile networks

RFC number	Title	Comment
IETF RFC 3261 [17]	Session Initiation Protocol	Commonly used protocol for initiation of communications sessions in packet switched technologies, used in 3GPP IMS mobile environment as well as in general Internet communications environments.
IETF RFC 5194 [i.35]	Framework for Real-Time Text over IP using the Session Initiation Protocol (SIP)	The document lists the essential requirements for real-time Text-over-IP (ToIP) and defines a framework for implementation of all required functions based on the Session Initiation Protocol (SIP) and the Real-Time Transport Protocol (RTP). This includes interworking between Text-over-IP and existing text telephony on the Public Switched Telephone Network (PSTN) and other networks. This standard refers to IETF RFC 4103 [20] for initiation, and transport of RTT.
IETF RFC 4103 [20]	RTP Payload for Text Conversation	This memo describes how to carry real-time text conversation session contents in RTP packets. Text conversation session contents are specified in Recommendation ITU-T T.140 [16]. One payload format is described for transmitting text on a separate RTP session dedicated for the transmission of text. This RTP payload description recommends a method to include redundant text from already transmitted packets in order to reduce the risk of text loss caused by packet loss.
IETF RFC 8373 [i.41]	Negotiating Human Language in Real-Time Communications	The document defines new Session Description Protocol (SDP) media- level attributes so that when establishing interactive communication sessions ("calls"), it is possible to negotiate (i.e. communicate and match) the caller's language and media needs with the capabilities of the called party. This is especially important for emergency calls, because it allows for a call to be handled by a call taker capable of communicating with the user or for a translator or relay operator to be bridged into the call during setup. However, this also applies to non-emergency calls
IETF RFC 9071 [25]	RTP-Mixer Formatting of Multiparty Real-Time Text	The document provides enhancements of real-time text (as specified in IETF RFC 4103 [20]) suitable for mixing in a centralized conference model, enabling source identification and rapidly interleaved transmission of text from different sources.

B.1.4 Standards and profiles In Mobile Multimedia Telephony from 3GPP, GSMA and ATIS

Standard	Title	Summary
ETSI TS 124 238 [i.25]	Session Initiation Protocol (SIP) based user configuration; Stage 3	The document provides a Session Initiation Protocol (SIP) based protocol framework that serves as a means of user configuration of supplementary services in the IP Multimedia (IM) Core Network (CN) subsystem. The protocol framework relies upon the contents of the Request-URI in a SIP INVITE request to enable basic configuration of services without requiring use of the Ut interface. The document is applicable to User Equipment (UE) and Application Servers (AS) which are intended to support user configuration of supplementary services.
ETSI TS 124 173 [i.24]	IMS Multimedia telephony communication service and supplementary services; Stage 3	The document specifies on a protocol level use of SIP for mobile conversational real-time multimedia communication and for specified Supplementary services. It specifies that real-time text, audio and video and data can be used together in a call with media as specified in ETSI TS 126 114 [8]. It also specifies that interworking may be done with other SIP based services by accepting communication with services not marking its communication with IMS tags.
ETSI TS 122 226 [i.19]	Technical Specification Group Services and System Aspects; Global Text Telephony (GTT); Stage 1	The TS contains the core requirements for the Global Text Telephony feature, which are sufficient to provide a complete feature to incorporate in conversational services. It defines the requirements for GTT to be understood as a framework to enable real-time transmission of text, for the purpose of a text based or text supported conversation between users. Text may be transported alone or in combination with other media in the session, especially video and voice. Thus the GTT enables text conversation to be included in any mobile conversational service. The term GTT is used as a common term for full functionality real-time text and for conversational text with lower functionality. When used over IP networks GTT fulfils real-time text functional requirements. The term GTT is used also in a number of other 3GPP specifications and can be seen as synonymous with real-time text and real-time text when used over packet networks. (Sometimes called GTT-IP). Annex A of the document covers total conversation: total conversation adds text conversation to multimedia protocols in a standardized way, so that simultaneous communication in video, text and voice is accomplished.
ETSI TS 123 226 [i.22]	Technical Specification Group Services and System Aspects; Global Text Telephony (GTT); Stage 3 (Release 13)	This is the architecture specification for Global Text Telephony (GTT), which is synonymous with real-time text (real-time text) when transported over packet networks. The document describes briefly GTT implementation in different technical environments, where use over packet networks is called GTT-IP. GTT-IP is originally detailed in ETSI TS 126 235 [i.63] by specifying codecs to use for real-time text, audio and video, but is now superseded by ETSI TS 126 114 [8] where up to date real-time text ,audio and video codecs and their use is specified for Multimedia Telephony.

Standard	Title	Summary
ETSI TS 122 173 [i.18]	Multimedia Telephony Service and supplementary services; Stage 1	This is the main service specification for mobile telephony and multimedia telephony, including total conversation. It says in clause 4.1: "*- IMS Multimedia Telephony is a service where speech, and speech combined with other media components, is the typical usage but the service is not limited to always include speech, it also caters for other media or combinations of media (e.g. text and video). ...This service specification says in its clause 4.2 that it fulfils the total conversation requirements:* *"IMS Multimedia Telephony service includes the following standardized media capabilities:* *- Full duplex speech;* *- Real-time video (simplex, full duplex), synchronized with speech if present;* *- real-time text communication;* *...* *The support of each of these media capabilities is optional for a UE. (= User Equipment; editor note)* *At least one common standardized format (e.g. JPEG, AMR) shall be supported per media type.* *NOTE: IMS Multimedia Telephony service fulfils the service requirement for the total conversation in Recommendation ITU-T F.703. The IMS Multimedia Telephony service shall support the following handling of media* *- Adding, removing and modifying individual media to/from an IMS Multimedia Telephony communication"* The document continues providing service level brief descriptions of supplementary services, of which Conference (CONF), Three-Party (3PTY), and Explicit Call Transfer may be of specific interest for real-time text and total conversation application.
ETSI TS 126 114 [8]	IP Multimedia Subsystem (IMS); Multimedia Telephony; Media handling and interaction	This is the specification for media handling in Multimedia Telephony. The document includes specification of how the three main conversational media; audio, video and **text (real-time text)**, are to be handled. The descriptions include the intended user experience, the transport protocol details, the addition of the media by the Session Description Protocol (SDP) to the communication, and quality aspects. The real-time text presentation is specified to use Recommendation ITU-T T.140 [16] and the transport is specified to use IETF RFC 4103 [20] updated by IETF RFC 9071 [25].
ETSI TS 122 101 [i.17]	3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Service aspects; Service principles	This is the highest level service specification of the 3GPP mobile communication system. It contains the following clauses of interest for total conversation and real-time text. *"7.2.2 IP multimedia (IM) sessions* *IP multimedia services are not the evolution of the circuit switched services but represent a new category of services, mobile user equipment, services capabilities, and user expectations. Any new multimedia service, which may have a similar name or functionality to a comparable standardized service, does not necessarily have to have the same look and feel from the user's perspective of the standardized service. Voice communications (IP telephony) is one example of real-time service that would be provided as an IP multimedia application.* *The following basic requirements are to be supported for IP multimedia:* *- IP multimedia session control shall be based on SIP.*

Standard	Title	Summary
		- All session scenarios shall be supported; i.e. Mobile Originating and Mobile Terminating sessions against Internet/Intranet, CS or IM Mobile, ISDN, PSTN call party. *- MSISDN and SIP URL numbering and addressing schemes shall be supported.* *- IP multimedia applications shall as a principle, not be standardized, allowing service provider specific variations.* *7.2.4 real-time text Conversation* *real-time text (real-time text) conversation is a service enabled in 3GPP networks by the Global Text Telephony (GTT).* *- GTT enables real-time, character by character, text conversation to be included in any conversational service, Circuit Switched as well as IP based.* *- It is possible to use the text component in a session together with other media components, especially video and voice.* *- Interworking with existing text telephony in PSTN as well as emerging forms of standardised text conversation in all networks is within the scope of this feature.* *- The text media component can be included initially in the session, or added at any stage during the session.* *- The text component is intended for human input and reading, and therefore supports human capabilities in text input speed. The character set support is suitable for the languages the users communicate in.* *- GTT specifies limited interoperation with Multimedia Messaging Services including a possibility to divert to messaging in case of call failure and sharing user interface equipment and external UE interfaces.* *Clause 10 specifies emergency communication with support for audio, video and real-time text, and thus total conversation.* *Clause 10.1 specifies general conditions valid for most types of emergency communications.* *Clause 10.4.2 IMS Multimedia Emergency Sessions is of most interest for real-time text and Total conversation because of its support for IP Multimedia Telephony. It starts:* *10.4.2.1 General* *For IMS emergency calls towards IP PSAPs, other media types may be supported by the UE and the IMS, subject to regulatory requirements.* *The media types that may be supported during an IMS MES include:* *- Real-time video (simplex, full duplex), synchronized with speech if present;* *- Session mode text based instant messaging;* *- File transfer;* *- Video clip sharing, picture sharing, audio clip sharing;* *- Voice; and* *- real-time text.* *Clause 10.8 Supplementary service interaction during emergency calls, has restrictions which are important to consider when planning language and modality conversion during emergency communications. The user is not enabled to invoke e.g. conference services during emergency calls. And it is of importance that it is the user device which initiates the emergency communications, so that e.g. addressing and location information provision is properly done as in any emergency communication".*

Standard	Title	Summary
ETSI TS 122 228 [i.20]	3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Service requirements for the Internet Protocol (IP) Multimedia core network Subsystem (IMS); Stage 1	This TS defines the service requirements from users' and operators' perspective for the support of IP multimedia applications through the IMS. General aspects of using services in the IMS packet switched technology are specified, such as addressing negotiation of features and media. Nothing specific is said about real-time text. real-time text is included in the IMS Multimedia Telephony concept specified on the service level in the document. Interworking with other IMS networks and technologies such as the Internet is briefly specified. WebRTC web based communications technologies are specified to be enabled to use IMS features in clause 11 of the document. This may be of interest for real-time text and total conversation use.

Standard	Title	Summary
ETSI TS 123 228 [i.23]	3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; IP Multimedia Subsystem (IMS); Stage 2	System aspects of IMS systems. This is the high level system specification for IMS, including how IMS Multimedia Telephony is handled. It is not media specific, so it handles all media including real-time text in the same way. Clause 7.5.2 has requirements of importance for real-time text and total conversation users. It requires a capability to negotiate languages and modalities in the communication. This feature can be used for invocation of relay services or routing to agents with matching language and modality capabilities. *"7.5.2 Negotiation at IM session invocation* *It shall be possible for the capability negotiation to take place at the time of the IP multimedia session invocation. Refer to clause 7.3 for further details on capability negotiation on IP multimedia session invocation.* *A UE should support negotiation of the user's desired language(s) (as defined in IANA) and modalities for spoken, signed and written languages.* *The system should be able to negotiate the user's desired language(s) and modalities, per media stream and/or session, in order of preference.* *A service provider shall be able to pass language and modality information between the endpoints. With respect to the language and modality information, there are no other service provider actions required."* *However the only language and modality specification mechanism mentioned in deeper technical 3GPP specifications is in ETSI TS 124 229, where only the lang attribute in the Session Description protocol is mentioned which is not specified for negotiation and without methods to assign priority to the different language and modality preferences declared. IETF RFC 8373 provides another mechanism, but it is not supported when the present document was authored.* *Interworking between voice, video and real-time text in WebRTC and Multimedia Telephony is specified in general in clause U.1.3.4 eIMS-AGW (IMS Access GateWay enhanced for WebRTC)* *The IMS-AGW enhanced for WebRTC (eIMS-AGW) is a standard IMS-AGW with the following additional mandatory characteristics and functions:* *- The eIMS-AGW may be used to perform any transcoding needed for audio and video codecs supported by the browser.* *- When GTT service is required, the eIMS-AGW shall perform transport level interworking between T.140 over Data Channels and other T.140 transport options supported by IMS.* *More specifically, real-time text, video and voice interworking between IMS Multimedia Telephony and WebRTC is specified in clause U.1.5.3 Protocol architecture for T.140.* *And U.1.5.4 Protocol architecture for Voice and Video".*

Standard	Title	Summary
GSMA NG.114 [i.33]	IMS Profile for Voice, Video and Messaging over 5GS V 5.0	The document defines a profile for voice, video, RCS Messaging and MSRP based Enriched Calling services over IMS, as well as SMS by listing a number of NG-RAN, 5GC, IMS core and UE features and procedures that are considered essential to launch interoperable services. The defined profile is compliant with and based on: 1. 3GPP specifications related to 5GS, voice and video services over IMS and SMS, and 2. GSMA specifications related to RCS Messaging and MSRP based Enriched Calling. The scope of this profile is the interface between the UE (User Equipment) and the network. The profile does not limit, by any means, deploying other standardized features or optional features, in addition to those defined in this profile. It specifies in clause B.2 that for real-time text (called GTT in this specification) "IETF RFC 4103 [20] shall be used as specified in ETSI TS 126 114 [8] when so required by regulation". By that it specifies a profile for total conversation
GSMA IR.94 [i.32]	IMS Profile for Conversational Video Service	The document defines an IMS profile by listing a number of Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Evolved Packet Core, IMS core, and User Equipment (UE) features which are considered essential to launch interoperable IMS based conversational video service. A video profile for IMS suitable for mobile total conversation together with the VoLTE concept.
ATIS 0700029 [i.1]	real-time text Mobile Device Behavior	This Standard specifies certain aspects of the mobile device behavior for handling real-time text (real-time text) to facilitate communication between mobile devices (including emergency services) across multiple Commercial Mobile Service Providers (CMSPs). This standard specifies details of the functionality of mobile end user devices with real-time text in USA. It is based on IETF RFC 4103 [20] and ETSI TS 126 114 [8].
ATIS 0700030 [i.2]	real-time text End-to-End Service Description Specification	This Standard defines the real-time text end-to-end service behavior for the handling of real-time text (real-time text) in support of the IP transition in order to facilitate a consistent use of real-time text across multiple Commercial Mobile Service Providers (CMSPs). This standard specifies details of the functionality of the mobile communication services with real-time text in USA. It is based on IETF RFC 4103 [20] and ETSI TS 126 114 [8].
GSMA IR.92 [i.31]	IMS Profile for Voice and SMS	The document defines a profile for voice over IMS over LTE, and for SMS over IP and SMS over NAS, by listing a number of Evolved Universal Terrestrial Radio Access Network (EUTRAN), Evolved Packet Core, IMS core, and UE features that are considered essential to launch interoperable services. The defined profile is compliant with 3GPP specifications. The scope of this profile is the interface between UE and network. This is the VoLTE concept for mobile real-time communication. It specifies in clause B.2 that for real-time text (called GTT in this specification) "IETF RFC 4103 [20] shall be used as specified in ETSI TS 126 114 [8] when so required by regulation".
GSMA 5G New Calling whitepaper [i.30]	GSMA Foundry 5G New Calling whitepaper 2023	This concept specifies an IMS communication concept with some new features and characteristics. It makes use of WebRTC technologies, which will mean IMS Data channel transport for real-time text. It refers to GSMA NG.114 [i.33] for technical details. This profile can be used for implementing total conversation.

B.1.5 Standards for Web Technologies

Standard	Title	Summary
WebRTC: Real-Time Communication in Browsers [28]	WebRTC: Real-Time Communication in Browsers	The document provides a standardized application program interface to create communications applications in web communication technologies. The main part describes how to use video audio and data channels in web technologies. It contains an accessibility annex where real-time text is specified (currently January 2023 out of date, but in progress of being updated).
W3C RAUR [i.57]	RTC Accessibility User Requirements	The document outlines various accessibility related user needs, requirements and scenarios for Real-Time Communication (RTC). These user needs should drive accessibility requirements in various related specifications and the overall architecture that enables it. It first introduces a definition of RTC as used throughout the document and outlines how RTC accessibility can support the needs of people with disabilities. It defines the term user needs as used throughout the document and then goes on to list a range of these user needs and their related requirements. Following that some quality related scenarios are outlined and finally a data table that maps the user needs contained in the document to related use case requirements found in other technical specifications. The document mentions real-time text, video, relay services, emergency services and user interface aspects.
IETF RFC 8865 [i.40] [i.40]	T.140 real-time text Conversation over WebRTC Data Channels	The document specifies how a Web Real-Time Communication (WebRTC) data channel can be used as a transport mechanism for real-time text using the ITU-T Protocol for multimedia application text conversation (Recommendation ITU-T T.140 [16]) and how the Session Description Protocol (SDP) offer/answer mechanism can be used to negotiate such a data channel, referred to as a T.140 data channel. By including real-time text in this way and video as specified in other WebRTC specifications, toral conversation can be handled in the WebRTC technology and used in the Internet as well as in 5G mobile networks.

B.1.6 Standards for relay service use

Standards	Title	Summary
ETSI ES 202 975 [4]	Human Factors (HF); Requirements for relay services	The document specifies requirements for relay services provided over ICT networks. It is intended to give information suitable for incorporation into contracts between commissioning agents and relay service providers. The document is applicable to all kinds of relay services which enable a user with functional limitations related to hearing, vision, speech or cognitive functions, or combinations thereof, to converse with other users. The document applies to text relay services, speech-to-speech relay services, video relay services, and captioned telephony services. Requirements are specified for services provided on a 24/7 basis, as well as for limited-hour services. The document does not place requirements on network operators. Communications assistants and sign language interpreters' "Communications Assistant" (CA) is defined for the purpose of the document as a person working in a relay service with media conversion, as a human intermediary; including sign language interpreters for video relay services.

Standards	Title	Summary
Recommendation ITU-T F.930 [i.46]	NON-TELEPHONE TELE-COMMUNICATION SERVICES Accessibility and human factors; Multimedia telecommunication relay services	Recommendation ITU-T F.930 [i.46] provides a functional description of four common types of relay services in use today: text relay, video relay, captioned telephone service relay and speech-to-speech relay. Additionally, it lays out specific functional requirements of relay services pertaining to equipment, call set-up, call experience, emergency communications and message retrieval. Telecommunications relay services enable persons who have hearing or speech disabilities and who otherwise would be unable to engage in voice telecommunications, to make voice telephone calls to other persons. In all forms of relay services, persons with disabilities connect to a communication assistant via a communications medium that is accessible to them. The communication assistant acts as an intermediary in the telephone call and converts between the accessible communication medium and voice, which is relayed from and to the person on the other end of the call.
NENA Video Relay Service & IP Relay Service PSAP Interaction Information Document [i.51]	National Emergency Number Association (NENA) Accessibility Committee NENA Video Relay Service & IP Relay Service PSAP Interaction Information Document	NENA Video Relay Service (VRS) and Internet Protocol Relay Service (IP Relay) PSAP Interaction Information Document is intended to provide guidelines for PSAPs and recommendations to the FCC regarding: Emergency calls to 9-1-1 via Video Relay and IP Relay Services (or similar third party entity) Effective relay of such calls to the appropriate Public Safety Answering Point (PSAP) Interaction between the caller, the Communication Assistants (CAs) and the PSAP Telecommunicators.
IETF RFC 9248 [i.42]	Interoperability Profile for Relay User Equipment	Video Relay Service (VRS) is a term used to describe a method by which a hearing person can communicate with a sign language speaker who is deaf, deafblind, or Hard of Hearing (HoH) or has a speech disability using an interpreter (i.e. a Communications Assistant (CA)) connected via a videophone to the sign language speaker and an audio telephone call to the hearing user. The CA interprets using sign language on the videophone link and voice on the telephone link. Often the interpreters may be employed by a company or agency, termed a "provider" in the document. The provider also provides a video service that allows users to connect video devices to their service and subsequently to CAs and other sign language speakers. It is desirable that the videophones used by the sign language speaker conform to a standard so that any device may be used with any provider and that direct video calls between sign language speakers work. The document describes the interface between a videophone and a provider.

B.1.7 Standards for emergency communications

Standards	Title	Summary
ETSI TS 103 479 [5]	Emergency Communications (EMTEL); Core elements for network independent access to emergency services	The core elements for network independent access to emergency services provide facilities that support mapping and routing functions for emergency communications. The document contains procedures for routing the call based on the caller's location and other conditions to the responsible emergency call centre. Other functional elements and necessary protocols and procedures enabling interoperable and secure implementations are also specified to allow multimedia communications as they evolve. These media include audio, video, and real-time text and by that total conversation. real-time text is specified to use IETF RFC 4103 [20] updated by IETF RFC 9071 [25].

Standards	Title	Summary
ETSI TS 101 470 [29]	Emergency Communications (EMTEL); total conversation Access to Emergency Services	The document defines conditions for using total conversation for emergency services with more media than in the regular voice call providing opportunities to more rapid, reliable and confidence-creating resolution of the emergency service cases compared to plain voice emergency calls, and especially for enabling access of emergency services or making them more usable for those persons who may have little or no use of voice telephony because of disabilities related to hearing, speech or other human communication functions.
ETSI TR 103 201 [i.12]	Emergency Communications (EMTEL); Total Conversation for emergency communications; implementation guidelines	The TR provides guidance for developers and PSAPs planning to implement total conversation for emergency communications, and for users of the total conversation service. The document covers emergency calls with the full media set of total conversation as well as subsets of the media, except voice calls in which no assisting service is needed. The document was intended to indicate gaps in standardization when it was created, and it has not been updated.
ETSI TS 103 478 [i.14]	Emergency Communications (EMTEL); Pan-European Mobile Emergency Application	The document presents the Pan-European Mobile Emergency Application (PEMEA) architecture provides the requirements and architecture for a solution to provide emergency application interconnection. It specifies the protocols and procedures enabling interoperable implementations of the architecture and provides extension points to enable new communication mechanisms as they evolve. This captures the case of total conversation and instant messaging.
ETSI TS 103 945 [i.16]	Emergency Communications (EMTEL); PEMEA Audio Video Extension	The document describes the PEMEA Audio Video (PAV) capability, and the need for this functionality. The required entities and actors are identified along with the protocol, specifying message exchanges between entities. The message formats are specified and procedural descriptions of expected behaviours under different conditions are detailed.
ETSI TS 103 871 [i.15]	Emergency Communications (EMTEL); PEMEA real-time text Extension	The Pan-European Mobile Emergency Application (PEMEA) architecture provides a framework to enable applications supporting emergency calling functionality to contact emergency services while roaming. PEMEA caters for a range of extension capabilities, including real-time text (real-time text) which provides a text based character by character exchange capability between the App user and the PSAP. The document provides a specification for a real-time text (real-time text) capability for PEMEA. Together with the audio video extension, it provides total conversation access to emergency communications [29].
ETSI TS 103 919 [6]	Emergency Communications (EMTEL) Accessibility and interoperability of emergency communications	Specifies requirements on emergency communications to be accessible and interoperable within Europe. Specifies requirements on the whole chain from user to PSAP. Intended as a pre-step for creation of an EN with title "Accessibility and interoperability of emergency communications to become harmonized".

Standards	Title	Summary
ETSI TS 123 167 [i.21]	Universal Mobile Telecommunications System (UMTS); LTE; IP Multimedia Subsystem (IMS) emergency sessions (ETSI TS 123 167 [i.21])	The document specifies system aspects of emergency communications from mobile users and call-back from emergency services. The document is valid for all three media of total conversation, and it has very little media specific statements. "*Clause 1.1 Architectural principles has two points about voice, video and real-time text (called GTT in this specification) :* *26. When a call is established with a PSAP that supports voice only, voice media is supported and GTT if required by local regulation or operator policy.* *27. When a call is established with a PSAP that supports voice and other media, voice, GTT and other media according to ETSI TS 122 101 (e.g. video, session mode text based instant messaging) can be used during an IMS emergency session if required by local regulation. This media may be used in addition to or instead of voice and/or GTT.*" Clause 4.5 Media, says "- *When the call is established with a PSAP that supports voice only, voice and subject to local regulation, GTT media is allowed during the IMS emergency session.* - *When the call is established with a PSAP that supports voice and other media, subject to UE and network support for the other media and local regulation, voice, GTT and other media according to ETSI TS 122 101 can be used during the IMS emergency session.* - *For sessions with a PSAP that supports voice and other media, media can be added, modified or removed during the IMS emergency session (e.g. adding video to a voice call) per media negotiation in ETSI TS 123 228.* - *When a PSAP that supports voice and other media attempts to add media, the media shall be added if accepted by the UE.*" Clause 7.5.2 mentions that it is possible to have emergency communications with a PSAP/Emergency centre connected via IP using SIP but no specific procedures are specified.

B.1.8 Use for automatic speech-to-text

Use of Automatic Speech to text has been developed for commercial use and the following are some examples available for assistive living:

- Solution that provides free app that captions phone calls using AI. It can convert text-to-speech and speech-to-text in real-time while remaining 100 % private. With this solution, the receiver of the call hears a natural-sounding voice, and whatever they say appears on the screen. One of the benefits of this app is that the recipient of the call does not have to install the app. This solution bridges the gap between the deaf and hard of hearing community and hearing people.

- Another solution is a speech-to-text app for deaf or hard of hearing individuals that uses captions to live transcription i.e. voice-to-text in real-time. It is ideal for group conversations and for one-on-one conversations.

- Other real-time transcribe solutions allow automatic speech transcriptions in near-real-time while offline. Some solutions allow text typing during real-time conversations and the ability to change the font size displayed.

- There are solutions that can convert speech to text and allow replies to messages also speech or text. The app can read out messages for the blind using a voice synthesiser. This app requires both parties to have it installed.
- Other solutions can support different languages and accents for hard of hearing individuals.

Annex C (informative): Simple video quality assessment

C.1 Introduction

This annex describes a simple method to assess video quality during real-time video communication corresponding to the requirements of EN 301 549 [1], clauses 6.5.1 to 6.5.4 and in clause 5.2.2 of the present document.

Modern networks and video communication devices usually easily surpass the video quality requirements specified in clause 5.2.2 of the present document, so a simple method as the one presented in the present annex can be sufficient for verifying the provided quality. If large deviations from the values for passing the tests are achieved, an evaluation with other means is recommended to be performed to assess if other sources than the ICT under test can be the source of the deviation.

NOTE: To suit the informational annex format, normative requirements in the present annex have been changed to "should".

C.2 Tools and test setup

1) Two pieces of user communication equipment A and B in stands in the same location are needed. They should be able to make calls with audio and video through a communications system with connections corresponding to what users will have.

2) A video camera which records video at at least 30 frames per second and audio. A mobile phone with tripod mount, and video recording app may be sufficient for this task, but it may be more convenient with more stationary equipment with zoom function so that it is easy to set a suitable view of the test equipment. A video resolution of VGA (640x480 pixels) is sufficient. Most video cameras perform 1080p resolution (1 920x1 080 pixels) which is more than sufficient.

3) A small device visualizing sound. It may be a specialized device, or a mobile phone with a suitable app. There is no need of any detailed visualization. It is sufficient with any kind of easily seen graphical change by a strong sound. The view of the sound should not be delayed more than 20 ms after the sound.

4) A video analysis program with functions for viewing video picture by picture. This is a normal function in many video editing programs. A video editing app in a mobile phone may be sufficient for this task, but it may be more convenient with equipment with a larger screen. A possibility to view the volume of sound in the video during single stepping can also be used instead of the device visualizing sound.

C.3 Tests

C.3.1 Resolution

Requirement: At least QVGA = 320x240 pixels are required and should be provided with sign language in view.

Set up the user equipment A and B a meter apart aiming in the same direction. Arrange the video camera so that its view of B nearly fills the view of the video camera.

Turn off audio in A, and set up a video call from A to B.

Start the video recording.

Stand in front of A with arms and hands stretched straight out to the sides of your body. Back off until your fingertips are at the edges of the picture. Move the right hand to beside your face, faced forwards and with fingers just slightly separated. After a second in this position, move the other hand faced forwards in circles about 50 cm wide perpendicular to a line to A, at least three turns about one second per turn.

Stop recording.

Analyse the video recording:

1) Check if the fingers of the right hand beside the face are seen clearly separated. while the left hand is still.
2) Check if the fingers of the right hand beside the face are seen clearly separated. while the left hand is moving.
3) Check if the fingers of the left hand are seen clearly separated. while the left hand is moving.

Passing check 1 is required by EN 301 549 [1], clause C.6.5.2.

Check 2 corresponds to the basic need in real use with sign language. Equipment with low performance can blur the picture when there is a lot of motion.

Check 3 corresponds to requirements for good sign language perception. Equipment with low performance can blur moving objects.

The theory behind this test is that at least three pixels in width are needed for each finger to be clearly seen. With the distance of the fingers from A, they would take up three pixels in width if the transmission conveys 320 pixels in width. The amount and speed of motion corresponds approximately to what is used in sign language. Modern communication equipment usually surpasses this requirement with good margin.

C.3.2 Frame rate

Requirement: At least 20 frames per second.

Analyse the video from clause C.3.1, where the left hand moves in a circle. Step through the video in single frame presentation during one second. Count how many times the left hand moves to a new position during that second.

Check if it is over 20. then the test passes.

Theory: The requirement is to transfer 20 pictures per second or over for reliable perception of the rapid movements occurring in sign language. It is common to send 30 pictures per second, but lower rates can appear either occasionally during heavy motion, or throughout the call when some resource is limited. (e.g. network bandwidth of equipment computing power or by exaggerated resolution setting)

C.3.3 Synchronization of video vs audio

Requirement: Video should be less than 100 ms before or after audio at the receiving side.

Setup:

Place A and B in different rooms. Turn on audio.

Place the audio visualizing device close to B facing in the same direction as the screen of B.

Place the video camera so that the screen of B and the sound visualizing device are captured.

Set up a call from A to B.

Start recording with the video camera.

Clap your hands with a distinct hand movement about one meter in front of A a few times.

Disconnect and stop recording.

Analyse the video with single frame stepping. Find the frame where the hands clap and make note of the time of that frame. Find the frame where the sound of the clap is presented on the sound visualizing device.

Evaluate the time difference between the image of the clap and the sound of the clap.

Check if audio is within 100 ms before or after video.

The test passes if the time is within these limits.

NOTE: Many video cameras record with 30 frames per second. Then the uncertainty of this measurement is only ±33 ms, which is a bit too high to be good for measuring a 90 ms asynchronism. If any uncertainty arises about pass or fail, then repeat the analysis for 4 different claps and take the mean of the results as the measured value.

Theory: For good benefit for lip-reading, The asynchronism should be within the range.

C.3.4 Video latency

Requirement. End-to-end latency of video: < 400 ms but better if is < 100 ms.

Note that this is not expressed as a requirement in EN 301 549 [1], because it depends on network conditions, but is of value to know. Networks are usually very rapid today.

Test setup alternative 1: Place A and B close to each other, both facing towards the video camera.

Turn audio off in A and B.

Set up a call between A and B.

Start video recording.

Place your hands so that they are seen by both A and the video camera on a distance of about 60 cm from both.

Clap your hands a few times with distinct motion.

Stop recording.

Analyse by single frame stepping in the video analysis program the time of the clap directly from the hands, and the time of the clap when it is presented in B, and evaluate the time difference as the resulting video latency.

Check if the latency is < 400 ms. Then the test passes.

Test setup alternative 2: Place a mirror in front of A.

Set up a call between B and A.

Place yourself so that your hands are seen by both B and the video camera.

Make video recording of B and the hands when they clap a few times in front of B.

Stop recording.

Analyse the time difference of the clap directly and when the mirrored image of A is presented in B. Divide the result by 2 to get the end-to-end latency.

Check if the latency is < 400 ms. Then the test passes.

Theory: If end-to-end latency is longer than 400 ms, sign language users may experience uncertainty if their expression was understood by the other party and their repetition may collide with the answer.

NOTE: The video latency and the video-audio asynchronism are useful for evaluation of the total conversation synchronism in EN 301 549 [1], clause 6.2.4, note 5 or clause C.6.7.

Annex D (informative): Change history

Date	Version	Information about changes
January 2024	V0.0.1	Initial version with Annex A and B only
January 2024	V0.0.2	EditHelp rescue and comments
February 2024	V0.0.3	Cleanup after ***editHelp!*** comments
September 2024	V0.0.4	First complete working draft of the ES.
October 2024	V0.0.5	Revised working draft of the ES - for public review associated with the workshop on total conversation interoperability and user equipment requirements
December 2024	V0.0.6	Completing the draft with clauses 7 and 8
March 2025	V0.0.7	Internal review and comments resolution
April 2025	V0.0.8	Final version
April 2025	V0.0.9	Final draft after EditHelp feedback
April 2025	V0.0.10	Final draft after clause header format corrections
May 2025	V0.0.11	Alignment of audio clauses because of changes in draft EN 301 549 and addition of a note detailing aspects of RTT presentation

History

Document history			
V1.1.1	June 2025	MAP process	MV 20250810: 2025-06-11 to 2025-08-11