ETSI TR 103 708 V1.1.1 (2022-08)

Human Factors (HF);
Real-Time Text (RTT) in Multiparty Conference Calling

Reference

DTR/HF-00103708

Keywords

accessibility, HF, ICT

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 Web server (https://ipr.etsi.org/).

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™** and **LTE™** 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.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Human Factors (HF).

Modal verbs terminology

In the present document "**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.

Executive summary

Real-Time Text (RTT) is text communication sent as it is created without any specific sending action by the user. It is required in EN 301 549 [i.13], including its use in multiparty calls, but has lacked details on user interface requirements and how to accomplish the multiparty function.

The present document explains the use and characteristics of RTT including its multiparty aspects. It also provides an analysis of documents from other standardisation bodies, and briefly proposes modifications to them to make lower layers of RTT implementation support multiparty calling and to ensure that emergency services are able to support RTT with multiparty functionality. Finally, detailed change proposals for EN 301 549 [i.13] version 3.2.1 are included in the final clause of the present document and in Annex A.

Introduction

The present document was developed to support future upgrading of those parts of the EN 301 549 "Accessibility requirements for ICT products and services" [i.13] standard that relate to the accessibility of Real-Time Text (RTT) when used in multiparty/conference applications. Multiparty usage is of particular importance in ensuring that communication with Emergency Services will meet the accessibility needs of persons for whom voice communication is undesirable or impossible. The inclusion of multiparty requirements will also ensure that EN 301 549 [i.13] remains relevant to the design of interoperable RTT and Total Conversation services.

1 Scope

The present document establishes user interface guidelines for RTT conference call interfaces and identifies technical support needed to implement the guidelines. The present document:

- introduces RTT;
- addresses the types of media and user interface elements that can be associated with RTT;
- describes some RTT use cases, highlights issues particularly relevant in multiparty scenarios;
- explains technical support issues related to text creation and presentation, and the standards and related services associated with RTT transport and control.

The final and crucial part of the present document proposes some potential changes and additions to EN 301 549 [i.13] to ensure that it covers the additional accessibility issues that arise when considering multiparty scenarios.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

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 are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI ES 202 975: "Human Factors (HF); Requirements for relay services".

[i.2] ETSI TS 101 470: "Emergency Communications (EMTEL); Total Conversation access to Emergency Services".

[i.3] ETSI TR 103 201: "Emergency Communications (EMTEL); Total Conversation for emergency communications; implementation guidelines".

[i.4] ETSI TS 103 478: "Emergency Communications (EMTEL); Pan-European Mobile Emergency Application".

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

[i.6] ETSI TS 123 167:"Universal Mobile Telecommunications System (UMTS); LTE; IP Multimedia Subsystem (IMS) emergency sessions (3GPP TS 23.167)".

[i.7] 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.8] ETSI TS 124 147: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Conferencing using the IP Multimedia (IM) Core Network (CN) subsystem; Stage 3 (3GPP TS 24.147)".

[i.9] 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)".

[i.10] ETSI TS 124 371: "Universal Mobile Telecommunications System (UMTS); LTE; Web Real-Time Communications (WebRTC) access to the IP Multimedia (IM) Core Network (CN) subsystem (IMS); Stage 3; Protocol specification (3GPP TS 24.371)".

[i.11] 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)".

[i.12] ETSI TS 126 236: "Universal Mobile Telecommunications System (UMTS); LTE; Packet switched conversational multimedia applications; Transport protocols (3GPP TS 26.236)".

[i.13] EN 301 549 (V3.2.1): "Accessibility requirements for ICT products and services" (jointly produced by ETSI/CEN/CENELEC).

[i.14] GSMA PRD IR.92: "IMS Profile for Voice and SMS".

[i.15] GSMA PRD IR.94: "IMS Profile for Conversational Video Service".

[i.16] GSMA PRD IR.51: "IMS Profile for Voice, Video and SMS over untrusted Wi-Fi access".

[i.17] GSMA NG.106: "IMS profile for Video, Voice and SMS over trusted Wi-Fi access".

[i.18] GSMA NG.114: "IMS Profile for Voice, Video and Messaging over 5GS".

[i.19] GSMA NG.115: "IMS Profile for Voice, Video and Messaging over Untrusted WLAN Connected to 5GC".

[i.20] IETF BCP 47: "Tags for identifying languages", M. Davis, A. Phillips, September 2009.

[i.21] IETF RFC 3261: "SIP: Session Initiation Protocol", J. Rosenberg et.al., 2005.

[i.22] IETF RFC 3550: "RTP: A Transport Protocol for Real-Time Applications", H. Schulzrinne et.al., 2003.

[i.23] IETF RFC 4103: "RTP Payload for Text Conversation", G. Hellstrom, P. Jones, 2005.

[i.24] IETF RFC 5194: "Framework for Real-Time Text over IP Using the Session Initiation Protocol (SIP)", A. Van der Wijk, G. Gybels, 2008.

[i.25] IETF RFC 6497: "BCP 47 Extension T - Transformed Content", M. Davis, A. Phillips, February 2012.

[i.26] IETF RFC 8373: "Negotiating Human Language in Real-Time Communications", Gellens R., 2018.

[i.27] IETF RFC 8825: "Overview: Real-Time Protocols for Browser-Based Applications", 2021.

[i.28] IETF RFC 8865: "T.140 Real-Time Text Conversation over WebRTC Data Channels", Holmberg C. and G. Hellström, 2021.

[i.29] IETF RFC 8866: "SDP Session Description Protocol", A. Begen et al., 2021.

[i.30] IETF RFC 9071: "RTP-Mixer Formatting of Multiparty Real-Time Text", DOI 10.17487/RFC9071, Hellström, G., (July 2021).

[i.31] IETF RFC 9248: "Interoperability Profile for Relay User Equipment", DOI 10.17487/RFC9248, Rosen, B., June 2022.

[i.32] ISO/IEC 10646:2020: "Information technology - Universal coded character set (UCS)".

[i.33] Recommendation ITU-T F.700 (2000):"Framework Recommendation for multimedia services".

NOTE: Available at https://www.itu.int/rec/T-REC-F.700-200011-I/en.

[i.34] Recommendation ITU-T T.140 (1988): "Protocol for multimedia application text conversation".

[i.35] Recommendation ITU-T V.18 (2000): "Operational and interworking requirements for DCEs operating in the text telephone mode".

[i.36] ETSI TS 103 871: "Emergency Communications (EMTEL); PEMEA Real-Time Text (RTT) Extension".

[i.37] W3C Recommendation 05 June 2018: "Web Content Accessibility Guidelines (WCAG) 2.1".

3 Definition of terms, symbols and abbreviations

3.1 Terms

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

communications assistant (CA): person providing alternative communication for a two-party conversation, multiparty meeting, or broadcast

NOTE: Communication assistants include captioners/transcribers, sign language interpreters, spoken language interpreters, relay service operators, call handlers, telephone operators, etc.

conference floor control: functionality for conference administrators and attendees to manage a formal or informal communication queue or other communication resources

NOTE: Examples of conference floor control include a "hand-raising" feature and a "request to share screen" feature.

continuous real-time conversation: type of organization in conversation and discourse where one participant's contribution is made available to other participants while it is being made

programmatically determinable: able to be read by software from developer-supplied data in a way that other software, including assistive technologies, can extract and present this information to users in different modalities

NOTE: WCAG 2.1 [i.37] uses "determined" where this definition uses "able to be read" (to avoid ambiguity with the word "determined").

Real-Time Text (RTT): form of a 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

relay service: electronic communications service that enables users of different modes of communication (e.g. text, sign or speech) to interact by providing conversion between different modes of communication, usually through a communications assistant

Total Conversation service: audiovisual conversation service providing bidirectional symmetric real-time transfer of motion video, text and voice between users in two or more locations (from ETSI TS 101 470 [i.2])

WebSocket: computer communications protocol, providing interaction between a web browser (or other client application) and a web server facilitating real-time data transfer from and to the server

3.2 Symbols

Void.

3.3 Abbreviations

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

3GPP	3G (mobile) Partnership Project
4G	4th Generation (mobile networks)
5G	5th Generation Mobile Networks
API	Application Programming Interface
BCP	Best Current Practice
CN	Core Network
EMTEL	Emergency Telecommunications
GSMA	GSM Association
HF	Human Factors
ICT	Information and Communication Technology
IETF	Internet Engineering Task Force
IM	Instant Messaging
IMS	IP Multimedia Subsystem
IP	Internet Protocol
IR	International Roaming Expert Group
ITU-T	International Telecommunication Union - Telecommunication standardization sector
LTE	3GPP Long Term Evolution (4G)
NG	Next Generation
PEMEA	Pan-European Mobile Emergency Application
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RFC	Request For Comment
RTP	Realtime Transport Protocol
RTT	Real-Time Text
SIP	Session Initiation Protocol
SMS	Short Message Service
TC	Technical Committee
UE	User Equipment
UI	User Interface
VoIP	Voice Over IP
WCAG	Web Content Accessibility Guidelines
WebRTC	Web Real Time Communication
WLAN	Wireless Local Access Network

4 Real-Time Text (RTT)

4.1 What is RTT?

RTT, or Real-Time Text, sends text characters input by a user shortly after they are entered. This contrasts with most text chat services where the user has toperform a confirmation step before the composed message is sent, usually by tapping a "Send" button. At the receiving end of an RTT conversation, the recipient sees the words from other participants as the characters are entered. The immediacy of the sending and receiving allows a dialogue that has a flow closely approximating that of a spoken dialogue.

Persons who use RTT use it for conversations in the same way as persons who can talk and hear use voice communications (or Voice and RTT together).

4.2 RTT is for conversation not messaging

Users may choose to communicate using a stand-alone messaging service or a chat feature in a conferencing system. They are making an active choice to use this less immediate, asynchronous form of communication.

The provision of a messaging/chat capability cannot be seen as providing an effective means for persons who cannot effectively make spoken contributions to effectively participate in a real-time person-to-person conversation or a conference call.

5 Multiparty calls

On a multiparty call (where most participants are using speech), including conference calls/meetings, the participants should not ideally talk at the same time, but in practice it occurs. If two or more participants speak at the same time it can rapidly lead to the multiparty call becoming ineffective for most of the participants. Multiple participants simultaneously using RTT could also have a similar effect, but the fact that text stays on the screen and the fact that text from different sources is presented with separation makes it feasible that some simultaneous text communication from different sources can be manageable. To what degree this is possible depends on the situation and how the conference is managed.

An important factor that leads to simultaneous contributions is a lack of awareness of when other participants have stopped contributing and when it is a suitable time for them to contribute. With voice communication, it is possible for most participants to hear when other participants are contributing, but they may not be able to identify who is speaking.

When several RTT users are in a multiparty call they need to be made aware when other RTT users are contributing. They also want to identify the other RTT users when they contribute. These are issues that are addressed in the present document. When contributions are being made both by voice and by RTT, the awareness of when someone is contributing becomes a much more complicated issue as contributions may be being made using a means that the user cannot, or is not able to, continuously monitor. This is even more true when video is used for information, e.g. by using sign language. User interface techniques that support users in having maximum awareness of when another person is actively contributing is a significant part of the scope of the present document.

6 User interface elements and use cases

6.1 Background

Availability of RTT is essential or desirable in several different situations in calls with two or more participants. This clause lists several valid use cases, with different combinations of media, different use of RTT, different characteristics of the devices for generation and presentation of RTT, and different means for management of the expressions in the different media in the call. The purpose is to be a base for checks that the specified characteristics of the user interfaces fulfil the user needs in these use cases.

6.2 Types of media

6.2.1 Real-Time Text (RTT)

The important characteristic of RTT is that text is transmitted at the same rate as it is produced, so that the receiver can follow the senders' thoughts as soon as they are turned into words. There will be no excess waiting time for completely expressed messages, similar to spoken communication in audio and signed communication in video. All three real-time continuous exchanges enable rapid interaction in conversation. A commonly used coding and presentation standard for RTT clarifying the characteristics of RTT is Recommendation ITU-T T.140 [i.34].

RTT may be used so that all call participants are enabled to create and send RTT to the other participants, and the text being presented in readable chunks growing in real time with indication of source. The presentation should provide an approximate view of the relative timing of text from different parties.

RTT may also be used by one or more participants in a call, where other participants communicate by speech or signing. In such situations, a human-transcribed or automatic translation may be included between the users of the different media in both directions.

A third way to include RTT in a call is to include translation from other media only for creation of RTT, while all participants have means for presentation of multiparty RTT, and those who need or prefer to express themselves in RTT are able to send RTT.

6.2.2 Audio

Audio is often used in calls with two or more participants to convey speech between the parties. The most common configuration is that audio presented to each participant is mixed from all other participants, except from those whose audio is muted. Transmission from each party can usually be muted both by the party and by a meeting administrator. Reception to a party can be controlled by the party and sometimes by the meeting administrator.

Audio is different from all other media described here in that the media itself does not occupy any screen area. However, small screen areas are needed for visual indication of audio, and for speaker identification.

Speech can be reliably perceived only from one participant at a time. Only extremely brief expressions from more than one party can be perceived, and then with reduced reliability. Therefore, multiparty calls are managed so that for most of the time only one participant at a time talks. The management may be self-controlled by the participants or made automatically or made by technical and human means by a meeting administrator.

Audio may also convey other sounds than speech, e.g. important environment sounds from an emergency scene in an emergency call.

6.2.3 Video

Video in calls with two or more participants is most often used for showing the participants, in order to aid in communication (e.g. facial expressions, body language) and facilitate camaraderie.

Sometimes the video view is used for conveying sign language.

Sometimes the video view is used to show a document, slideshow, or other resource, but presenting these components separately is becoming more common.

Video for conversational use is rarely used alone, but usually combined with other conversational media.

6.2.4 Total Conversation

Total Conversation enables real-time interaction between two or more participants in a call with the opportunity to shift focus immediately between communication modalities in the three real-time conversational media available in the service.

6.2.5 Messaging

Multiparty calling services commonly provide ways for the participants to send messages to each other. Sometimes these messaging functions work independently of any real-time session, sometimes messages can only be sent while there is a real-time call. All users or specific users can be addressed by the messages and answers. Some services offer only plain text messages, while others can include images, files and video clips in the messaging channel. In some services translation can be requested.

6.2.6 Documents, drawings and stored or streamed media

It is very common that documents and drawings are presented in multiparty conferences as well as stored or streamed video, audio and text media. The transmission of this media is usually controlled by a meeting administrator or a presenting participant. The layout on the screen is usually controlled by the receiving participant, deciding how large the visual media is and to what degree other visual components in the user interface are combined on the screen.

On small screens, users would benefit from the ability to temporarily hide selected visual user interface components while watching meeting presentations. When a user interface component is hidden, it may also be desirable to get indications when there is something new in the hidden component.

6.3 Use of media and user interface elements

6.3.1 Potential user interface components

A multiparty multimedia call may contain many user interface components which compete for presentation space and the awareness, perception, creation, and management by the users. It is therefore relevant to consider all commonly appearing user interface components in a multiparty call, when designing the user interface for RTT.

6.3.2 User generated media presentation

Audio is unique in that it requires no screen space for its media contents presentation. All the other features listed in clause 6.2 require presentation on a screen for them to be perceived by sighted participants, and each feature will compete with the others for screen space. Participants will only be able to view contributions from other participants that are using video or text presentation if they currently have a part of the screen dedicated to displaying that media.

Video from each participant is often presented in subsections of the screen. The selection, size and placement of the video views of the participants can usually be controlled by the receiving user, and by a meeting administrator. Sometimes the size, style, or placement of the view of a participant is also influenced by sound or other activity from the participant, with the intention that the current speaker will be shown with a higher priority than the other participants.

6.3.3 Active and waiting user activity indication

Hearing participants will immediately be aware when another participant is speaking. Hard-of-hearing or Deaf participants may not be able to detect when one of the other participants is speaking, but often they are able to infer who is speaking by interpreting on-screen markers or placement of the participant information about the active speaker.

By default, users will only be aware that another participant is actively attempting to communicate with other participants if they are already monitoring the media channel that the participant is using, or if they infer it by interpreting on-screen markers or placement of the active speaker.

Knowing that someone is actively communicating is very important, knowing the means by which they are doing so is of secondary importance. It is, therefore, important that any means of identifying that a person is actively communicating should be available to all contributors at all times, irrespective of the means that they have chosen to use, or are only able to use. This will minimise the risk that someone using one communication method is not noticed by the other participants that are using a different communication method. As well as indication that a person is actively communicating, it is also important for other participants to be aware who is actively communicating at any time. The means of identifying that communication is taking place and the means of identifying who is communicating can often be the same.

It is more complex to automatically identify that video contains active communication than it is for text and audio, but it would be equally useful in an unmanaged conference with users mainly focusing on different media.

In some managed conference systems, there are controls that allow a participant to indicate that they wish to participate next (e.g. a "hand raising" system). This is sometimes supplemented by a system that allows other participants to view a queue of upcoming speakers. To ensure inclusive participation, it is important that all users, regardless of their chosen communication media type, can perceive the list of upcoming speakers.

6.3.4 User control of presentation and input

The sections above make it evident that there are many factors in a multiparty call that a user may need to control and be made aware of in an easily perceivable way. All the media types described in clause 6.2, except the audio media itself, may compete for screen space. In addition, the indications described in clause 6.3.3 may compete with each other and with the media presentation for the user's attention and screen space.

It is unlikely that a single interface design can present all the content from these competing sources, while satisfying all users in all situations. Users will need an easily handled means to select and configure the size and placement of what to present on the screen whilst simultaneously being aware that there are other contents available for presentation.

Controls will also be necessary to allow a participant to select which function (e.g RTT or text messages) will get text input, and which external devices will be used for audio, video, text and other input.

In addition to the fundamental complexity of presenting the correct content and status information at the right time, there will be a need for participants to check back and review the earlier RTT and message contributions. One reason is that not all participants will be able to reliably switch and refocus their attention to follow the flow of the conversations in real time.

6.4 Use cases

6.4.1 Use case variations

The following clauses contain descriptions of use cases that include RTT. The variations in use of RTT and other media are wide. The use case list can therefore not be comprehensive, but instead is a collection with sufficient variation to serve as examples for brief checking that a real or tentative implementation of RTT will be accessible and usable.

RTT is often indicated as a media of interest for communication with and between Deaf and hard-of-hearing persons and with persons with speech related disabilities, providing them with the same responsive way of communicating that is available to those able to use voice communication. The rapidity and lack of delays in RTT can also be of value in a written conversation between any users. Therefore, most of the use cases contain users with accessibility needs.

6.4.2 Call using RTT within a small group of Deaf persons

Four Deaf persons engage in a multiparty call with video, audio and RTT. Three of them use mobile phones, and one makes use of a small computer. They all prefer to use sign language, and therefore select a view where the three video images of the other participants take up most space in the user interface. They get into discussing a medicine, and in order to provide the name with exact spelling in a way that the others can keep, the person talking about the medicine types its name in RTT. The computer user gets the RTT field presented immediately and sees the name being typed. The ones with mobile phones get an indication in the user interface that there is new text in RTT to read, so they switch the view to one with place for the RTT content and can make note of the name before they return to the view with larger video images.

6.4.3 Deaf person calling emergency service and using RTT

A Deaf woman with a mobile phone gets into an emergency situation, so she calls the emergency number with RTT activated. The country she is in supports calls with RTT and audio to the emergency service. This is verified by a service discovery mechanism. The callers' language preferences are conveyed to the emergency service, so the call is routed to a suitable call taker. The call taker answers and carries out an initial conversation in RTT. It is decided to add a poison expert to the call, so the call is extended to a three-party call with RTT, and the poison expert takes over the main discussion, but the call-taker stays on the call observing the outcome by following the RTT flow, and then helps to dispatch an ambulance. During the call, audio is used so that the poison expert can listen to the Deaf woman's breathing and the expert can confirm the diagnosis. During the call, all parties see the text flow from the parties as it is typed, and the text gets presented in readable chunks and has brief labels indicating the role of the emergency service participants and the name of the woman in the emergency situation.

6.4.4 Hard-of-hearing user talking with hearing friends

A small group of friends have started sending text messages within the group discussing how to get together. One man in the group is hard-of-hearing. The alternatives and opportunities to arrange how to meet up get complicated so one of the parties takes the initiative to start a real-time multimedia call with audio, video and RTT, so that they can resolve more rapidly what is possible and what is not. By hearing and seeing the others talking, even the hard-of-hearing person manages quite well to understand what is said. But occasionally it gets hard to understand, and then he asks, "please type that!". The friends are used to this form of communication, so they divert to typing the hard part in the RTT connection. The hard-of-hearing man understands already what it was about by seeing the first half in text of what he missed in voice, and can say "thanks", and the call can continue mainly by speech.

The user interfaces contain the video images of the friends, and the one speaking gets presented with the largest video image. There is also a small area available for typing and reading RTT. The participant who is typing is indicated by a label first in an RTT field growing in real time as text is received. If occasionally more than one is typing, they get one field each with a leading label. Fields with already completed RTT text chunks are closed and move up and eventually out of the currently presented part of the RTT.

6.4.5 Deaf user participating in conference getting transcription support

A few Deaf persons participate in a large remote conference with predominantly hearing persons. There are audio, video, messaging, RTT, and document presentation components available. The meeting organizer has made it possible to subscribe to a real-time transcription of what the speaker says in the conference. Also, questions by the audience are transcribed. The transcription is sent by RTT so that the Deaf participants can get the contents in real-time. Thereby they can also raise their hands and get the floor to contribute to the discussion.

Some of the Deaf participants contribute by speech, while others type in RTT. The transcription is provided by automatic speech-to-text technology in RTT monitored by human text interpreters ready to jump in and transcribe when the quality of the automatic transcription is not sufficient.

All participants are notified when there is new RTT content, and all participants are able to read RTT.

6.4.6 Deaf user participating in conference contributing by text-to-speech

In this case, the situation is similar to the previous one. In this conference no visual presentations occur, so it is feasible for hearing users to participate with a telephone only. The Deaf participants are provided with speech-to-text and text-to-speech support in RTT, so that the hearing participants can perceive the contributions from the Deaf participants even with just a telephone. The text in RTT is spoken out in suitable chunks so that it is perceivable but is not delayed too much.

6.4.7 Deaf-Blind user participating in remote meeting

Deaf-Blind persons have wide variations in how they send and receive remote meeting contents. Some prefer hearing speech even with reduced hearing, some prefer seeing enlarged text with special contrast, and some prefer to read text through a braille display through screen-reader support.

This Deaf-Blind person prefers to get most input from the media in the call through a braille display and screen-reader support. The conference system has features for provision of automatic speech-to-text transcription of the speech part of the meeting to be sent by RTT to the Deaf-Blind participant. Others are also allowed to contribute directly by RTT.

Some meeting services may also provide integration with human transcription and sign language interpretation support.

The screen-reader support and presentation on the refreshable braille display has functionality for indication when there is new text available and for navigating the braille display throughout the Real-Time Text conversation.

The Deaf-blind person's own contributions are made by a regular keyboard or some accessible input method and can be directed towards RTT or text messaging depending on the urgency.

6.4.8 Person in a critical situation making an emergency call by RTT

A hearing man gets caught up in a dangerous situation. He wants to make an emergency call without speaking to avoid being discovered.

He calls the emergency number and sets up an RTT call with the emergency services. The call-taker answers and after a short dialogue, connects to the first responder in a three-party call with RTT. The first responder takes over the dialogue in text, and the call-taker stays on the call, seeing everything that the others type.

By using RTT they can minimize how much needs to be written before it is understood and thereby speed up how quickly the situation is handled.

6.4.9 Person in remote group meeting in occasional noise

Anne is a project manager in a food production company. She is sitting in an airport gate area and has a small remote group meeting with colleagues by her small computer. They have audio, video, RTT, chat and screen sharing available. They mainly communicate by voice. Anne explains the next step in a project, so that the colleagues can continue work while she is on the flight. The airport environment suddenly becomes very noisy, so she turns her microphone off and changes to typing in RTT. Text flows immediately, so the group can keep focus on the topic and continue the urgent preparation. Also, colleagues wanting to contribute realize that Anne may not hear what they say, so they also type in RTT when they want to comment. Eventually the noise goes away, and the group can return to mainly using voice.

6.4.10 Relay service using multiparty technology

A Deaf woman Anne gets an incoming voice call to her mobile phone. It is apparent by details in the call that no RTT is included and that the calling phone does not have RTT capabilities enabled. Anne realizes the need to have a text relay service included in the call. She answers by an operation that activates a three-party bridge in her phone which makes an outgoing call with RTT and voice to a preconfigured address of a relay service. The communications assistant answers and translates between voice and RTT in that call. The embedded multiparty bridge conveys only the voice part from the relay service operator to the incoming call, and RTT to Anne.

In this way, Anne can handle incoming voice calls to their own phone number.

In another situation Anne makes an outgoing call to a voice user by a simple operation that is expanded to first call the text relay service with RTT and voice and then call out to the voice user with voice enabled. The three-party bridge in her phone connects the suitable media of these calls so that the communications assistant can perform the relaying action between RTT and voice for Anne.

6.4.11 Using an RTT relay service to connect to a voice conference call

Peter, a Deaf person, wants to participate in a digital meeting with a majority of hearing participants. The meeting is accessed by a web address. Peter connects to the meeting through his web browser. Peter does not want to disturb the meeting by requiring RTT to be provided by the meeting organizers. Instead he also connects to an RTT relay service and provides meeting details so that they can connect to the meeting by web communication. When the meeting starts, the communications assistant types what is said in the meeting to Peter, and speaks Peter's RTT responses to the other meeting participants. Peter and the relay service operator can view documents and speaker list and he can operate the hand raising functionality directly in his own view of the meeting.

7 RTT text creation and presentation

7.1 Text creation and transmission

7.1.1 Variation in text creation means

RTT is intended to satisfy needs to communicate in a real-time conversational mode by text. That implies that parties in a call who want to send text need to be provided with suitable means to create text and direct it to the RTT transmission functionality in the ICT used for communication.

What is considered a suitable means to create text may differ between different persons and situations. ICT should be designed to enable such variation of text input means. Examples of means which may be supported are:

- On-screen keyboard for the language the user is familiar with and intends to use.
- Built-in physical keyboard for the language the user is familiar with and intends to use.
- External physical keyboard for the language the user is familiar with and intends to use.
- Automatic speech-to-text function supporting the language used in the call.
- Using the ICT device's copy and paste operations to input text as an occasional complement to other means.

- Braille keyboard, built in or external.
- Handwriting or fingerspelling to text conversion.
- Automatic sign language to text functionality.
- Text prediction methods used in combination with any of the other methods.
- Other methods for accessible text input by built-in or external devices.

7.1.2 Verification of produced text

The sending user usually needs a way to view the text that they input and approximately how it appears in relation to the timing of the text from other participants. Therefore, both currently created text and earlier created text should be presented in similar ways to how received text is presented.

7.1.3 Erasure

During text creation, and especially just after a text chunk is closed, it is not uncommon that the user discovers a mistake or wants to change a small part of text just sent. Therefore, erasure (usually a character at a time) of the latest piece of text sent should be possible, even if this means erasing back over a transmitted new line or other indicator of a completed chunk. Otherwise, frustration appears together with the awkwardness of explaining what was sent unintentionally.

Erasure should have the same effect on the presentation to the receivers as to the sender, so that they have an opportunity to have a common impression of the communicated information in the call.

Voice communication users do not have the option to erase and correct what was said, but are instead used to indicating and correcting mistakes quite rapidly by speech. Similar corrections in RTT takes much longer time, so therefore the erasure opportunity is provided.

7.1.4 Input language and character set

The users need to have the means to send text in the languages and character sets suitable for the calls they participate in.

Text has different writing directions in different languages. Support of a language implies support of the common writing direction of that language.

7.1.5 Sending of emoji

ICT users have been accustomed to use emojis in text communication. Transmitted text should use Unicode coding and support emojis from the Unicode set of emoji.

7.1.6 Graphic rendition

Rich text formatting features are seldom used in practice with RTT, but some RTT presentation standards allow text to be transmitted with a wide range of possible formatting applied. It is important that where the receiving RTT software does not support the formatting applied to the incoming text, it can be handled without the software entering an error state, distorting the displayed text, or otherwise disrupting the presentation.

Users can often, through their ICT system or app-specific settings, set the colours they prefer for reading RTT. These settings can be critical to the user for accessibility and readability reasons. Implementations of RTT presentation standards that support modifying the foreground and background colour of sent text can cause problems for the receiving user by overriding their carefully chosen colour preferences.

7.1.7 Time from text entry to presentation

Text characters should be transmitted as soon as they are created but may be stored, for a short while, for time-sampled transmission in small groups when so decided for transmission efficiency reasons. The maximum waiting time before transmission should be 500 ms because receiving users tend to experience slight annoyance and start to wonder when text will arrive if a piece of text is delayed more than one second after creation, and the receiving users may start repeating questions and causing confusion if display of received text is delayed more than two seconds after creation. The 500 ms waiting time allows for 500 ms network transmission time, which is suitable in many cases. Even 1½ seconds network transmission time can then be occasionally allowed before the transmission is experienced to be severely delayed, see Recommendation ITU-T F.700 [i.33].

The 500 ms waiting time is just a maximum to enable an acceptable conversational quality while causing acceptable network load and should not be seen as a design goal. A value of 300 ms is often used in the transport implementations for RTT, providing a more pleasant, smooth presentation and a more rapid real-time impression.

The RTT user relies on short waiting times before transmission in unmanaged meetings with mixed media. The time slot during which it is feasible to respond to a speaker is often quite short. Comments are better accepted if they come during the brief pauses that speakers make. Any delay risks making the comment appear to be late and irrelevant to what the speaker is currently talking about, or may be disturbing because the speaker has moved on to another topic.

When the multiparty call is arranged by a central conference bridge, there is a risk that the transport mechanism causes delays before text from the participants is sent to the receiving parties. If text transmissions from the users simultaneously sending text are distributed in time with e.g. 300 ms from each user, then it is possible to apply a rule that the bridge normally should send received text on to the receivers without any inserted waiting time. However, in order to not cause network congestion, the bridge should have a restriction for e.g. how many text transmissions per second it is allowed to send to each participant. Occasionally, many RTT users may happen to send text simultaneously causing the bridge to be forced to apply waiting time on text from some of the participants. Just as with voice in multiparty calls, the times are rare when many users contribute at the same time, so with good transport design, such moments with inserted delays will be very rare.

7.1.8 Support for user to send text when the situation allows

Different calls will have different requirements on users regarding sending while others are signing, speaking or real-time texting. In some cases, the users only need to observe and assess when it is appropriate to send some text, while in other calls, formal requests of the floor by hand-raising and waiting for their turn are required. For both these cases, it is essential that the user who wants to send RTT can observe the behaviour of the other participants and behave according to the conventions in each call.

7.1.9 Reliability versus rapidity in transport of text

The transport protocol of text should make it possible to meet both the rapidity expectations described above in clause 7.1.7 and reliability expectations expressed as quality level T2 in Recommendation ITU-T F.700 [i.33] about quality of conversational text even in a situation of some network communication problems. The measures for reliability should guarantee that no text is duplicated during error recovery because duplication can cause confusion among participants. Instead, it may be allowable that text is lost in rare failure situations in order to not hold up the communication in lengthy retransmission efforts. If transport functions find that text may have been lost, an accessible indication should be provided to the receiving user.

7.2 RTT text presentation

7.2.1 RTT text presentation basics

The display of text from the members of the conversation should be arranged so that the text from each participant is clearly readable, and its source and the approximate relative timing of entered text is represented in the display. Mechanisms for looking back through the contents of the current session should be provided. The text should be displayed as soon as it is received.

Different possible ways to arrange text for visual, tactile and audible presentation can be envisioned. The following clauses provide the most used ways depending on the medium.

7.2.2 Visual presentation variants

7.2.2.1 Introduction

These examples are presented as potential user interface solutions. No specific, prescriptive design is required by the present document or the related documents. Users should be provided with means to customize size, style and colours to some degree.

7.2.2.2 One column view

One potential way of multiparty RTT presentation shows all text from all participants in one column. Text from different parties is labelled with a readable identity. Text from each user is collected in readable chunks, delimited by an end of phrase or sentence, new line, or long inactivity. Text chunks delimited in that way are closed and flow chronologically with other closed chunks of text. Newly created or received text should be rendered chronologically more recently and predominantly than closed chunks. If multiple participants are creating text simultaneously, received text is added in multiple insertion points in the presentation area.

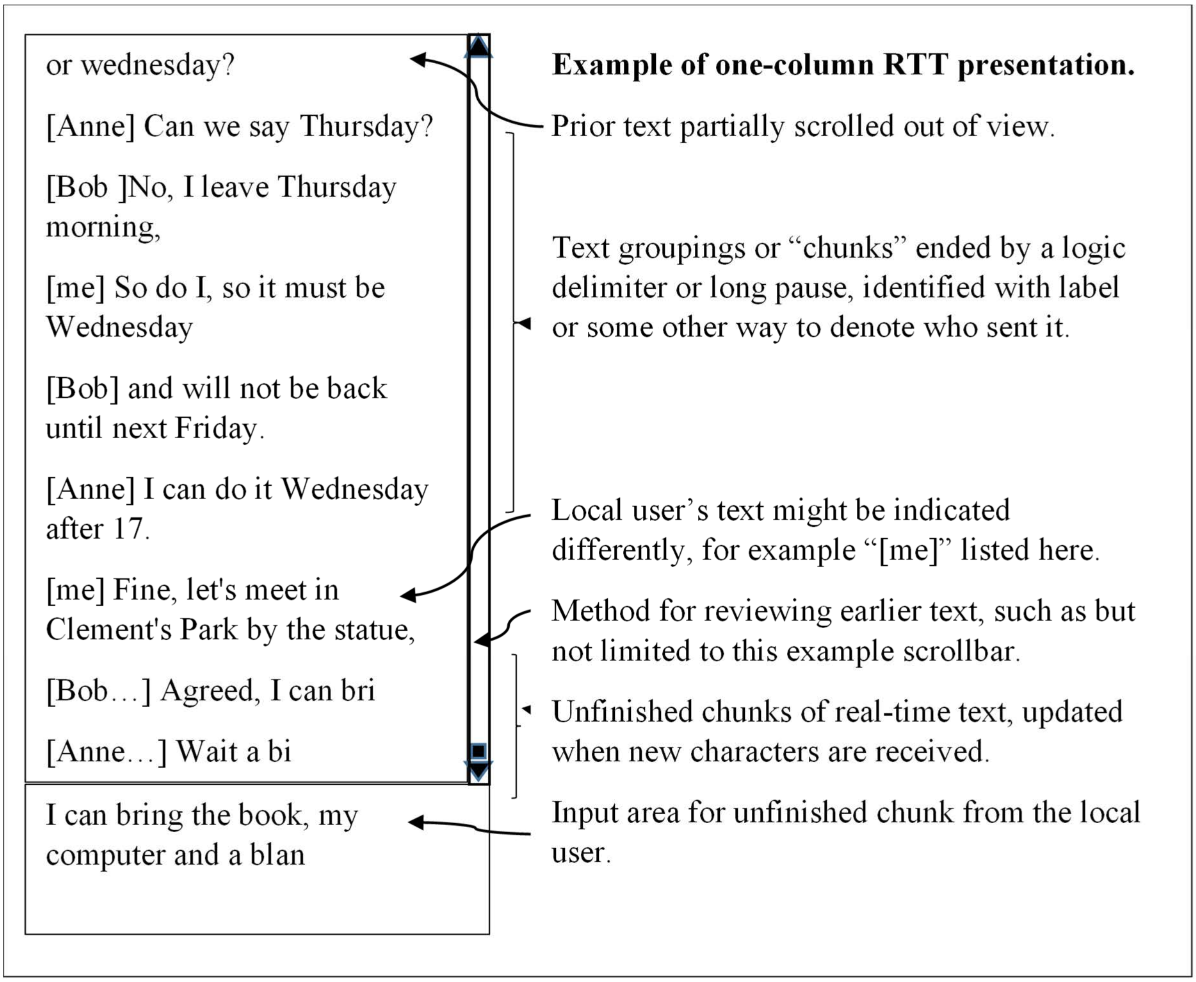


Figure 1: Example of a one-column presentation with explanations

7.2.2.3 One column per contributing user

Another potential way of presenting multiparty RTT is one column per contributing user. Each column is labelled with text representing the source. The vertical position of text indicates the approximate relative time it was completed.

7.2.2.4 Text beneath a video image of the corresponding participant

In calls where video has importance, it may be suitable to present RTT in a limited space under the video image of the source. As it is not often possible for the system to know when the presence of video is important, activation and deactivation of this way of presenting text beneath a video image should be under the control of the user.

7.2.3 RTT text presentation by screen reader with spoken output

When using a screen reader with spoken output for RTT contents, it is usually beneficial to not automatically read out received contents until a space or punctuation mark has been received or after some time of inactivity, so that complete words can be read. When the user has scrolled back to read earlier received text, the arrival of new text should not interrupt the current speech, though some implementations may choose to use unobtrusive audio cues to inform the user that new text has arrived.

7.2.4 RTT presentation by screen reader software with braille output

7.2.4.1 Notes on braille device user interfaces

7.2.4.1.1 Braille Devices

Braille devices typically include *braille keyboards* and *braille displays*. Many braille displays include an integrated braille keyboard. Some can act as standalone note taking devices, while some can be paired with another computing device such as a phone or laptop.

Some braille implementations may choose to set aside positions on the display to indicate when new text has arrived and other important events during the call. There should also be a method to toggle the single row of braille cells between the local user's created text and the text of other participants.

Rather than give specific user interface examples that may be misunderstood as prescriptive design advice, the goal of this clause is to explain some types of braille input and output, and how assistive technology implementors might integrate RTT with braille.

7.2.4.1.2 Braille Keyboards

The term "*braille keyboard*" is usually limited to devices that have chorded input for text entry, but no output (no refreshable braille cells). Braille keyboards feature 6 or 8 keys for chorded typing input, along with 1 or 2 keys that act as a spacebar. There may be other additional function buttons.

Figure 2: Portable 6-dot *braille keyboard* next to a modern smartphone for scale comparison.

7.2.4.1.3 Braille Displays

The term "*braille display*" is used to describe devices that have refreshable braille cells for *output*, but these devices may also contain a chorded braille keyboard for text input. The following figures show the most common style of single-row braille display, with braille cells in a single row. Buttons on the top and sides allow chorded typing and other control functions.

Figure 3: Two 40-cell braille displays with integrated 8-dot braille keyboards and additional buttons

Figure 4: Portable braille displays with 12 and 18 cells. Some have as few as 8 cells

Braille displays that do not include braille keyboards are most often used alongside a standard computer keyboard for text entry.

Figure 5: A braille display without a braille keyboard, used alongside a laptop keyboard for text entry

7.2.4.1.4 Braille Matrix Displays

Some newer consumer braille devices include a grid of refreshable pins that can render braille and other tactile graphics. For the purposes of the present document, these are referred to as "*braille matrix displays*." User interface conventions on these new devices are emerging, innovating, and likely to continue changing.

Figure 6: Two recent matrix-style braille displays that can render both braille characters and other tactile graphics

7.2.4.2 Anatomy of a Braille Display

Figure 7: Common features of braille displays. Not all features will be available on all displays, but the device pictured represents a common feature set for desktop braille displays

Each braille cell consists of moveable pins used to represent information, usually in the form of braille letters and other characters. Routing buttons above each braille cell allow contextual interaction with the information represented on the adjacent braille cell. For example, pressing a routing button over a link could activate that link, or pressing a routing button over editable text may position the text insertion cursor.

Panning buttons to either side of the cell array allow the user to panning the displayed text forward and back in various ways, usually one full row at a time.

Braille keyboard keys (usually 8 across the top) correspond to each of the 8 finger positions, and each pin of an 8-dot braille cell. In many western language braille tables, pressing the left index finger results in dot 1 (⠁) or the "a" character, while pressing both index fingers simultaneously results in dots 1,4 (⠉), which is the "c" character in most contexts. A more detailed explanation of braille context is outside the scope of the present document.

In addition to typing, the braille keyboard keys can be used to trigger "command chords" or a combination of dots that is not commonly used while typing. For example, simultaneously pressing the index, middle, and pinky fingers of the left hand results in dots 1,2,7 ("⡃ "), a "chord" that is not commonly used while typing braille characters. Command chords and other function buttons can be assigned to trigger features of the braille display or screen reader.

7.2.4.3 Example RTT Implementations with Braille Displays

7.2.4.3.1 The purpose of the following examples

Braille displays and the screen readers that support them vary greatly in terms of feature sets and user interface conventions, so the present document makes no specific recommendations about user interface when implementing RTT with braille.

The following are examples of *existing* implementations, and speculative interface ideas on newer matrix devices. These examples are presented for informative purposes only. These are not intended as recommendations or requirements.

7.2.4.3.2 Example: RTT using Reserved Status Cells

One braille implementation on a major mobile operating system utilizes a set of 3 reserved cells, which represent various states in a variety of contexts, including RTT. The user can choose whether the status cells are positioned on the right or left of the display. All other available cells on the row are used for rendered text.

During an RTT call, the reserved status cells can represent states such as whether there are new characters, and whether another RTT participant has completed typing. Pressing the routing button over the RTT status cells toggles the user's display between their own text, and the text of other participants. When the local user has selected to display the text of another participant who is actively typing, the text appears in near real time on the recipient user's braille display. Optionally, a physical button or a command chord (multiple braille keys pressed simultaneously) can be assigned to toggle the display between relevant areas of interest.

Figure 8: RTT implementation example showing cells reserved for status and text, with toggle functions assigned to a routing button and a braille chord

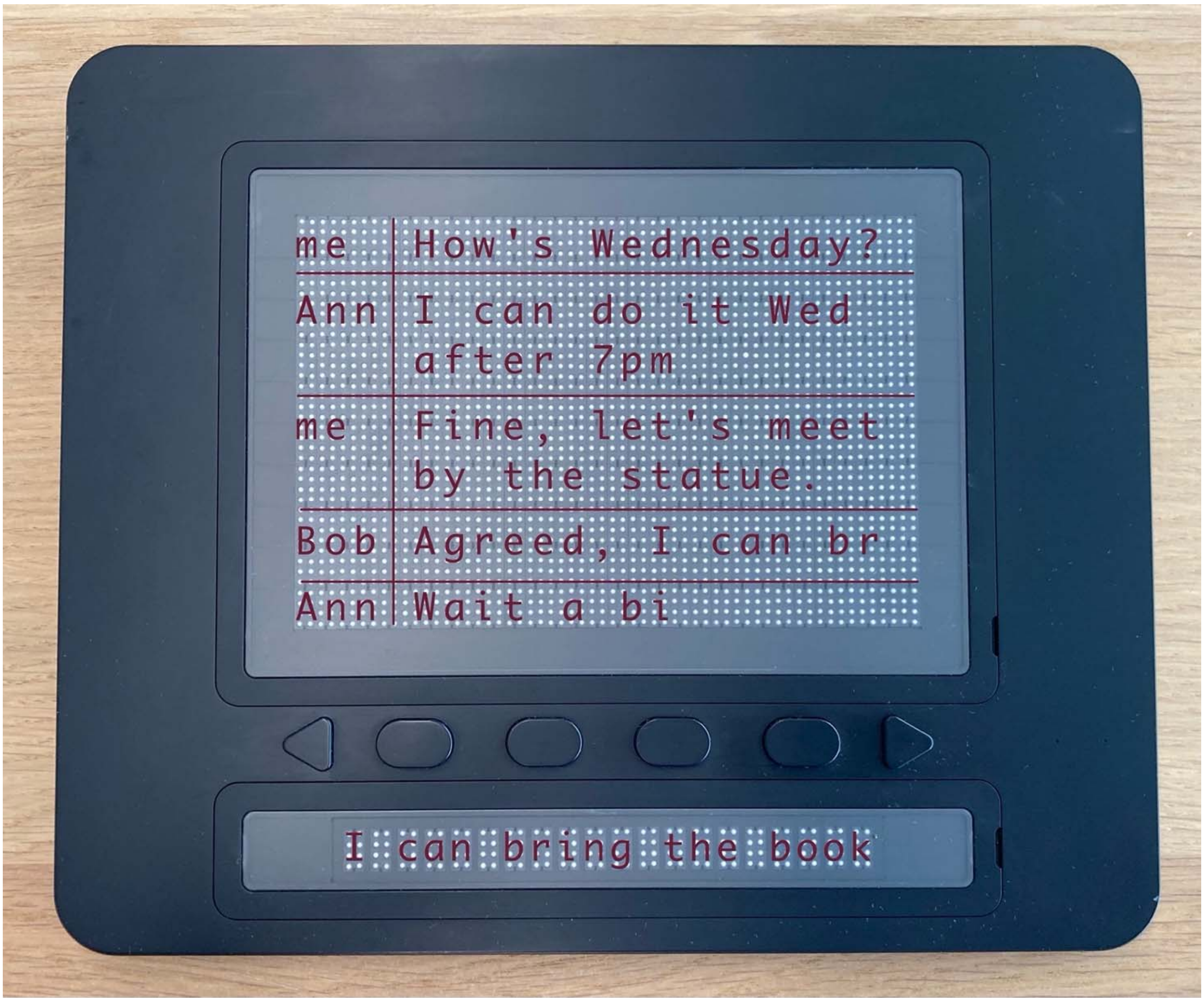
7.2.4.3.3 Example: RTT without Status Cells on Portable Displays

Portable displays may have as few as 8 braille cells. Some of the popular portable displays use 12 cells. In all cases, space is at a premium on these smaller displays, so many users choose to disable status cells to maximize the available braille real estate for text.

Toggling between the user's text entry and received messages can be assigned to a function key, or to a chorded braille command. On this existing implementation, the status of the RTT call states can be conveyed through notifications, which may display transiently or flash the display cells.

7.2.4.3.4 Example: RTT on a Braille Matrix Display

At the time of this writing, there are no known implementations of RTT on the newer *braille matrix displays*, but the following is presented as an example of how existing graphical user interface conventions for RTT may work on a two-dimensional tactile interface.



NOTE: Braille characters replaced with letter glyphs for understandability and demonstration purposes.

Figure 9: Possible RTT conversation layout on a braille matrix display

8 Control of RTT in multiparty calls

8.1 General aspects of control of RTT

As seen in clause 6, about user interface elements and use cases, many events and pieces of content compete for screen space and user awareness during a multiparty call with RTT. Convenient management of these is essential to ensure that the user feels in control of the call.

8.2 Control and indications of the visible user interface components

8.2.1 Control of the screen space

There is always a need to select which parts of a multiparty call gets screen space, and a need for ways to control the selection and placement of parts which are selected to be visible.

8.2.2 Scrolling in RTT presentation

The simplest example is the RTT display, where the whole call cannot be presented, resulting in older text chunks being scrolled out of the presentation area. When the user wants to review older parts of the call, it should be possible to scroll back in the RTT text. While in the scrolled position, arrival of new text should not make the presentation move, but instead an indication should appear to show that new text is available. The user should then have a way to easily scroll to a position where new text is automatically presented when it arrives.

8.2.3 Searching and selecting the RTT contents

It may be of value for the user to be able to perform a search within the RTT text of the current call.

When there are many contributing parties to RTT, then there may be situations, with small presentation areas, when users want to filter text and only have text from specific sources presented.

8.2.4 Varying the RTT presentation

Users should be provided with a means to vary characteristics of the RTT presentation for comfortable reading. For example, implementations could allow the user to adjust the placement and size of the RTT presentation area, size of text, user interface colours, etc.

8.2.5 Notifications and status information about participants and media

The user interface of multiparty calls should convey status information and notifications of events in the call. Events could be presented as a summary for the call, and/or as indications per participant. Some implementations present this information in the participant list. Examples of information that may be useful for implementations to convey:

- There is new RTT text available since the user last viewed the RTT presentation area.
- An RTT text chunk is completed.
- A user is currently speaking.
- There is new text chat message available since the user last viewed the text chat presentation area.
- The status of media production and reception capabilities for each participant, including mute state.
- The state of the RTT media reception path from each participant.
- Raised hands.
- Document or screen sharing in the presentation area.

When video is used in the call, it may also be helpful to provide the same information close to the video image of the person it belongs to.

8.2.6 Selection between meeting views

It is helpful for implementations to provide view configuration options for users. For example, some participants may prefer to use one or more of the following views within the course of a single meeting:

- large meeting presentation area and smaller views for other components;
- all video images presented along with RTT;
- RTT and messages, but no video images;
- RTT area with adjustable size;
- RTT and text messages combined into the same area.

8.2.7 Notifications by audio and visual means

Audible and visual notification should be possible to indicate when a new RTT text chunk is completed, and when a new chat message is received.

8.3 Control via assistive technologies

Users of assistive technologies, such as screen readers and braille displays, have modes of presentation, where they get summary notifications about status and events in the user interface. This can lead the user to move focus to the user interface element where more information is provided. Such notifications are available in the spoken or braille interface of screen readers. Developers of RTT and Total Conversation clients should ensure that their application user interface works well with assistive technology features.

9 Control of calls with RTT

There may be calls where most participants are not familiar with RTT, but one or a few users are depending on use of RTT for input or output or both. That situation can provide equal opportunities for RTT users to participate in several ways. E.g.:

- The RTT user or the meeting organiser connects a relay service to the call, which converts spoken information in the call to RTT towards the RTT user. The relay service can also read out RTT text from the RTT user to the other participants. For this to work, the RTT user and the relay service needs RTT support and access to all other media in the call, while the other participants do not need to handle the RTT part. The RTT user may need to be enabled to set up the call with the relay service.

- The RTT user can hear but not speak, and can therefore follow the contributions to the meeting by the others. For contributions by the RTT user, the user may type in the RTT area, and all other users may get RTT presented, or get a notification that there is RTT text to view. The meeting system may have a combined area for RTT text and messages so that all participants feel reasonably familiar with activating the view of these presentations. The users need to be provided with means to enable RTT for all in the call.

10 Transport of RTT with multiparty contents in various technical platforms

10.1 General transport considerations

Different meeting technologies are designed around different media transport mechanisms. It is therefore natural that transport of multiparty RTT needs to have different transport mechanisms in these different meeting technologies. Even if meeting systems may use any RTT implementation which meets the usability and accessibility requirements, it is beneficial if different meeting systems in the same meeting technology use the same RTT transport mechanism. That will ease interoperability both with other meeting systems in the same meeting technology and with meeting systems in other meeting technologies. Requirements for interoperability in RTT with emergency services also influence the selection of transport technology for RTT in the meeting systems.

Some transport mechanisms are already standardised for multiparty RTT. Others may need to be specified when multiparty RTT is about to be introduced in a multiparty meeting technology. This clause presents some already specified multiparty RTT transport mechanisms.

10.2 Multiparty RTT transport in centralized SIP conferences

For the IETF RFC 3261 [i.21] SIP call control protocol environment, the most commonly used transport standard for RTT is IETF RFC 4103 [i.23]. As with many other media transport protocols, IETF RFC 4103 [i.23] is based on the Real-Time Protocol RTP specified in IETF RFC 3550 [i.22]. RTP has specified ways to support media transport in multiparty calls which are suitable for RTT transport by IETF RFC 4103 [i.23], but until July 2021 there was no specific method specified, and therefore a risk that future implementations of bridges and terminals would be made to support different methods. Therefore, an update to IETF RFC 4103 [i.23] for multiparty calls was standardized and an "RTP-Mixer Formatting of Multiparty Real-Time Text" is available as IETF RFC 9071 [i.30].

IETF RFC 9071 [i.30] specifies a negotiation to check whether an endpoint is capable of using its method for transport of multiparty RTT. If it is, then a bridge can apply that method. If not, then a procedure is specified which provides a fallback method for multiparty RTT presentation in multiparty-unaware endpoints.

The method is suitable for use in SIP conferences with centralized control, which is used both in general VoIP services and in the mobile IMS area, specified by 3GPP and GSMA.

10.3 Multiparty RTT transport in WebRTC

A rapidly emerging real-time communication environment is WebRTC [i.27], where the intention is that endpoints are using standardized functionality in web browsers for their communication. RTT including multiparty transport for the WebRTC environment is specified in IETF RFC 8865 [i.28].

10.4 Multiparty RTT transport in PSTN

In the Public Switched Telephone Networks (PSTN), various text telephony protocols are specified for transmission of text in real-time with lower functionality than that specified for RTT. Different protocols are, or were, used in different countries, specified in Recommendation ITU-T V.18 [i.35]. They are in the process of being phased out, or are in some cases already abandoned but if there is any interest in having multiparty text calls in real-time in the PSTN, then it is possible with reduced functionality. The functionality limitations differ between the protocols. None of them can natively present text received intermixed from different sources in a readable way, so a multiparty mixer for conference-unaware endpoints similar to what is specified in IETF RFC 9071 [i.30], sections 4.2 and 6.1, about gateways to PSTN is needed for the multiparty case. With this method, the mixer transmits text in real time only from one source at a time, storing text from other sources in the mixer until a suitable moment for switching source is detected by the mixer. Each time a source switch is made, a readable label is inserted by the mixer indicating next source to get its text sent.

More information on interoperability can be found in IETF RFC 5194 [i.24], however without the multiparty aspects.

Some protocols transmit text more slowly than rapid typing, and some allow text transmission in only one direction at a time. No legacy protocol enables efficient simultaneous transmission of text and voice, while most protocols can be used for alternating transmission of voice and text by human operational procedures. Some legacy protocols have limited character sets, such as not handling any difference between upper and lower case. These limitations cause severe complications and a risk of delays when used for interoperability with full functionality RTT implementations. The protocol used in some Nordic countries, specified in Recommendation ITU-T V.18 [i.35], annex F, has sufficient speed and two-way simultaneous transmission so that some RTT-like communication could be achieved. But no implementation can handle fully functional multiparty RTT. The application would be very limited and implementation of this reduced functionality text communication with PSTN should only be recommended in regions where requirements for such fall-back interoperability still prevail.

10.5 Multiparty RTT transport in PEMEA

In the Pan-European Mobile Emergency Application (PEMEA) ETSI TS 103 478 [i.4], RTT is provided as an extension capability. There is work in progress on the RTT extension in draft ETSI TS 103 871 [i.36]. RTT in that specification is based on and mapped to the requirements in Recommendation ITU-T T.140 [i.34], and supports multiparty and control character sets. Each text fragment may contain up to 500 ms worth of typed characters from a specific sender. If a user's connection drops and then re-establishes, the transport is able to request all messages from a certain point onwards. This ensures that there is no loss or duplication of messages. Connections in this protocol are made using a secure WebSocket protocol. So far, this protocol is intended for use with emergency communications but can be adapted for general communications.

11 Interaction with services

11.1 Services of specific importance for RTT users

There are electronic communication services in which support for multiparty RTT will be of particular benefit for RTT users. This clause explains the reasons why multiparty RTT capabilities are so beneficial for RTT users when using these services.

11.2 Continuous real-time conversational services

Continuous real-time conversational services allow calls between user devices to be set up and have a specified combination of proposals for the audio, video and Real-Time Text conveyed between the participants in the call. The address used for call setup may be in any form established by the service provider e.g. a number in a number plan, or an address in an IP format user@domain, or just a username in a specific service.

There are several ways to include users in a call; calling out to a list of invited participants, by adding another party on request to an existing call, or by all participants calling in to a multiparty conference bridge.

By providing RTT in such calls, users who prefer RTT can use it in one or both directions and ask the other participants to use the different media in the directions that suit everybody's capabilities. If it is not feasible to ask everybody to use RTT directly, the user who prefers RTT may instead invoke a human operated or automatic transcription service to or from other modalities available in the call. Such services are called Relay Services and are described in ETSI ES 202 975 [i.1].

11.3 Conference services

Conference services are similar to multiparty continuous real-time conversational services described above. Conference services also usually support presentation of images and documents in the conference and provide text message-based chat functions. For an opportunity to interact in conferences, the rapidity of RTT is essential, and that all other participants become aware of new entries. An alternative is that the RTT users have support from a relay service that rapidly transforms the RTT words to spoken language. This is described in more detail in ETSI ES 202 975 [i.1].

Users of assistive technologies may need accessibility API support in the meeting application, including an accessible turn-taking UI, if they are to be able to fully participate in multiparty conference call.

11.4 Emergency services

Emergency services require multiparty calling in the media supported by the service. The person who first takes the emergency call often needs to make an assisted transfer of the call to some other agency for planning the action that results from the emergency call. The person who takes the call needs to stay on the call to just observe the communication until the call is progressing well with the next agency involved. For an RTT user, calling RTT supporting emergency services, both the bridge and the parties involved need to support multiparty RTT calling.

11.5 Relay services

11.5.1 The purpose of Relay Services

When there are one or more participants in a call who requires RTT in either or both directions, and it is not feasible to require all participants to create or read RTT in a corresponding way, then inclusion of a transcribing or transforming service can close the communication gap by transcribing or transforming language manually or automatically between different modalities.

Transforming or transcription services are often called relay services. This is especially the case when they are provided with support by national authorities to make electronic communications services accessible. Other cases are sometimes called remote interpreting services or text interpreting or caption services. To ease nomenclature, all variants are called relay services in the present document.

11.5.2 Functionality and connection alternatives

Requirements on relay services are specified in ETSI ES 202 975 [i.1]. However, the version valid at the time of writing the present document does not cover all requirements from modern communication situations.

A standard for the lower layers communication of user devices accessing both relay services and real-time conversational services is IETF RFC 9248 [i.31]. The present document includes specification on how to support multiparty calls including RTT mainly in the SIP call control environment based on IETF RFC 4103 [i.23] and IETF RFC 9071 [i.30], but also briefly on how to do it in WebRTC based on IETF RFC 8865 [i.28].

Relay services are named according to their main media or modality. Video relay services mainly handle conversion between sign language and speech, while text relay services handle conversion between RTT and speech. RTT is a valuable addition in video relay services, e.g. for cases when exact spelling is required. The relay services covered here are mainly text relay services.

There are many situations where multiparty calling relates to relay service implementation:

- The relay service itself is a three-party call even for a call between two persons. The three-party call with application-specific media mixing is established between the two call participants and a transforming entity. The bridge is often placed in the relay service, but placing it in the users' equipment can have some benefits.

- Relay service support is needed in multiparty meetings and conference services to enable participation by persons needing RTT communication in meetings where users predominantly use voice. The relay service will be needed when the users of the different media are not prepared to create media of the other kind, i.e. hearing, talking users are not prepared to type in RTT, and RTT users are not prepared or not capable of talking. RTT may be provided to all, but it is only essential that it is sent to the RTT users.

- Relay service support in meetings and conference services can also be implemented by the person who needs RTT connecting to the conference with one leg of a separate conference bridge and connecting to the relay service with the other leg. The relay service transforms between the modalities so that RTT does not need to be distributed in the conference. When the conference has shared documents and other conference features, it is usually expected that the relay service also has a link to the conference service for these features.

11.5.3 Delay caused by translation or interpretation

Translation or interpretation by relay services naturally takes some time even when the media are transmitted in real time. A delay is introduced, depending on the type and means of conversion. Automatic voice to text conversion is usually very rapid, while manual services can cause more delays. Conference participants may need to be made aware that some participants communicate with support of relay services and may have a slight delay in their interactions.

11.5.4 Queueing situations

When intending to set up a call with relay service support, a delay can appear before the call can be handled while waiting for a free communications assistant in the relay service. During the delay, the call is put in a queue. In that situation, accessible information about the queue situation should be provided to the users in voice and RTT. This is further elaborated in ETSI ES 202 975 [i.1].

11.5.5 Language and translation

The standard for the use of language tags to indicate language is IETF BCP47 [i.20]. The present document can be used to support decisions about engaging relay services in calls and conferences. An extension of BCP47 [i.20]; IETF RFC 6497 [i.25] specifies if media is transformed or translated by adding a T-extension to a language tag. It can be used in various places where language can be specified for call or media stream contents. Such places are in call or media declaration in IETF RFC 8866 "lang" attribute [i.29] where the language of call or media can be declared, or by IETF RFC 8373 [i.26] where the language in media can be negotiated. A user who has the same contents available in multiple modalities, e.g. spoken and RTT may benefit from knowing if a language stream is transcribed or not. For example, English produced by any form of transcription or translation from English in another form could get the language tag "en-t-en", while English generally is indicated by the tag "en".

12 Relations to specifications from other groups

12.1 General about support from other standardisation bodies

For support of multiparty calling with RTT, each implementation environment requires support for user interface features described in the present document, but also support from call setup and media transport functions in the lower layers of the target systems. This clause contains advice and observations about what modifications may be needed in the standards created by other relevant groups. The other groups are of course best at judging where and which modifications are needed for multiparty RTT, so the information in the following clauses should only be seen as brief hints and advice and a starting point for assessments from experts in the standardisation bodies.

Most standards and specifications about RTT refer to Recommendation ITU-T T.140 [i.34] for presentation, coding and general media session handling. Use of this common reference for the higher layers simplifies interoperability between technologies using different standards for media establishment and transport.

12.2 Relations to 3GPP specifications

One important source of specifications for electronic communications in mobile environments is 3GPP. RTT is already well specified in 3GPP specifications but, as in some other RTT specification collections, there is a lack of detail about how multiparty calling should be established and performed.

3GPP specification modifications are almost certain to be needed to ensure the success of attempts to specify multiparty RTT calling at the user interface level in the mobile communications area.

Although it is the responsibility of 3GPP groups and experts to fully assess and decide where and how to make modifications, some starting points for possible analysis and actions in 3GPP are proposed in this clause.

For traditional SIP based interactive communication, the main 3GPP specification covering media handling and interaction in multimedia telephony is ETSI TS 126 114 [i.11]. It contains technical details on how to use audio, video and RTT in interactive calls, providing a specification for Total Conversation and subsets thereof.

RTT is specified to use Recommendation ITU-T T.140 [i.34] on the presentation level and IETF RFC 4103 [i.23] on the transport level. IETF has recently (July 2021) published an update for IETF RFC 4103 [i.23] about how to use IETF RFC 4103 [i.23] in multiparty calling. The update that specifies RTP-mixer formatting of multiparty Real-Time Text is IETF RFC 9071 [i.30]. See clause 10.2.

A way to update ETSI TS 126 114 [i.11] to support RTT multiparty calling would be to add IETF RFC 9071 [i.30] for multiparty calling to where IETF RFC 4103 [i.23] is mentioned, and show the negotiation for multiparty RTT support during call establishment specified in IETF RFC 9071 [i.30] in the corresponding sections of ETSI TS 126 114 [i.11].

Similar modifications could be done in ETSI TS 123 226 [i.7] and ETSI TS 126 236 [i.12]. However, these specifications do not seem to be maintained, so it will be for 3GPP experts to assess if it is meaningful to make amendments to them.

In ETSI TS 124 147 "Conferencing using the IP Multimedia (IM) Core Network (CN) subsystem; Stage 3" [i.8], a paragraph could be added in clause 6.3.2 about support of the text (Real-Time Text) media for multiparty sessions in conference scenarios. However, this level of detail is not shown for other media, so it may be better to concentrate the details to ETSI TS 126 114 [i.11].

For WebRTC based real-time communication, establishment and transport of RTT is specified in IETF RFC 8865 [i.28], as described in clause 10.3 of the present document. IETF RFC 8865 [i.28] already supports multiparty calling, and IETF RFC 8865 [i.28] is already referred from ETSI TS 124 371 [i.10]. The user device is expected to have ETSI TS 126 114 [i.11] implemented which allows for WebRTC data channels. This might be sufficient for specification of WebRTC based Real-Time Text with multiparty support in 3GPP specifications.

There may also be a need for amendments to 3GPP emergency service specifications of access to IP based emergency services, starting with an analysis of ETSI TS 123 167 [i.6]. 3GPP experts are trusted to find the correct places and wordings.

12.3 Relations to GSMA specifications

12.3.1 Introduction

GSMA has published a number of specifications about the most widespread mobile systems and services. A few of them specify ways to implement conversational services in GSMA systems and services. These may benefit from analysis and modifications for adding specification of how multiparty RTT calling could be included in GSMA services. Target GSMA documents include:

- NG.114 "IMS Profile for Voice, Video and Messaging over 5GS" [i.18];
- IR.92 "IMS Profile for Voice and SMS" [i.14];
- IR.94 "IMS Profile for Conversational Video Service" [i.15];
- IR.51 "IMS Profile for Voice, Video and SMS over untrusted Wi-Fi access" [i.16];
- NG.106 "IMS profile for Video, Voice and SMS over trusted Wi-Fi access" [i.17];
- NG.115 "IMS Profile for Voice, Video and Messaging over Untrusted WLAN Connected to 5GC" [i.19].

GSMA NG.114 [i.18] is for mobile 5G networks, while the other referenced documents are profile variants for the same service where the User Equipment (UE) is in different network types:

- LTE;
- untrusted Wi-Fi™;
- trusted WLAN connected to 5G;
- untrusted WLAN connected to 5G.

NG.114 [i.18] and IR.92 [i.14] contain RTT.

These are all profiles specifying how to use 3GPP specifications, mainly ETSI TS 126 114 [i.11] for Voice, Video, RTT and Messaging over IMS networks. Profiling here means specifying more exactly how optional items from the referenced 3GPP specifications should be selected. That provides good opportunities for interoperability between devices and operators.

The following clauses provide brief descriptions of these documents. They also contain proposals for how to make the documents contribute to implementation of multiparty RTT calling in their specific areas, and how application of the present document in GSMA services can contribute to user satisfaction. The editing proposals are provided in italics.

12.3.2 Analysis of GSMA NG.114, IR.92 and IR.94

NG.114 [i.18] is titled "IMS Profile for Voice, Video and Messaging over 5GS".

IR.92 [i.14] is titled "IMS Profile for Voice and SMS".

IR.94 [i.15] is titled ""IMS Profile for Conversational Video Service".

These specifications are profiles for how to implement services with conversational voice and RTT over two different mobile network technologies. NG.114 [i.18] is for 5G networks and IR.92 [i.14] for 4G LTE networks. NG.114 [i.18] also specifies messaging and conversational video. IR.92 [i.14] also specifies IP based SMS and has a closely related document, IR.94 [i.15] specifying conversational video in LTE networks. NG.114 [i.18] and the combination of IR.92 [i.14] and IR.94 [i.15] thus specify profiles for Total Conversation (the combination of conversational voice, video and Real-Time Text) in the applicable networks. These profiles all refer to ETSI TS 126 114 [i.11], where it is specified how voice, video and RTT should be initiated, coded and transported in conversational mobile services based on IMS.

RTT is specified in annex B.2 of both NG.114 [i.18] and IR.92 [i.14], referring to Recommendation ITU-T T.140 [i.34] and ETSI TS 126 114 [i.11] for RTT. ETSI TS 126 114 [i.11] in turn refers to IETF RFC 4103 [i.23] for initiation and transport of RTT.

Annex B.2 of both NG.114 [i.18] and IR.92 [i.14] also add requirements beyond what is specified in ETSI TS 126 114 [i.11] on how to initiate and carry RTT in the GSMA services, especially for the network bearers and quality of service aspects.

In the current (as of December 2021) versions of GSMA NG.114 [i.18] and IR.92 [i.14], nothing specific is specified for how to handle multiparty RTT calling. Since a few supplementary services in IMS are forms of multiparty calling, section 2.3 of NG.114 [i.18] and IR.92 [i.14] can be consulted for a general specification for the handling of supplementary services, and then specifically section 2.3.3 "Ad-Hoc Multi Party Conference" for the most common form of multiparty calling. This section should be valid also for multiparty calls including RTT. Therefore, it seems appropriate to mention Annex B.2 in section 2.3.3.

To assure interoperability and good efficiency in multiparty situations, the transport standard IETF RFC 4103 [i.23] for RTT has recently (July 2021) acquired an update detailing RTP-mixer formatting of multiparty Real-Time Text in IETF RFC 9071 [i.30].

This update is not yet referenced in ETSI TS 126 114 (version 17.2) [i.11] which is the base for RTT calling for NG.114 [i.18] and IR.92 [i.14]. Therefore, it is recommended to add such references directly to these specifications.

A minimum would be to add references to IETF RFC 9071 [i.30] at suitable points in Annex B.2 of IR.92 [i.14] and NG.114 [i.18].

IR.94 [i.15] specifies how to add video to IR.92 [i.14], thereby providing the full set of media for Total Conversation in LTE, similar to what NG.114 [i.18] specifies for 5G networks.

12.3.3 Other GSMA profiles for conversational voice, video and RTT

The documents:

- GSMA IR.51 "IMS Profile for Voice, Video and SMS over untrusted Wi-Fi access" [i.16];
- GSMA NG.106 "IMS profile for Video, Voice and SMS over trusted Wi-Fi access" [i.17]; and
- GSMA NG.115 "IMS Profile for Voice, Video and Messaging over Untrusted WLAN Connected to 5GC" [i.19];

all detail how to handle conversational calls and messaging over various forms of Wi-Fi connecting to IMS networks. They follow the structure of IR.92 [i.14] and NG.114 [i.18], and specify how the situation differs from the ones specified in IR.92 [i.14] vs NG.114 [i.18]. RTT is not mentioned. Access over Wi-Fi is relevant also for RTT users.

It is recommended to add an annex, similar to Annex B.2 in IR.92 [i.14] and NG.114 [i.18], after the update for multiparty calling proposed above, but adjusted so that only QoS aspects of the Wi-Fi transport are included and not the mobile network aspects reflected in Annex B.2 of IR.92 [i.14] and NG.114 [i.18]. Insertion of links from the sections on ad-hoc conferencing to the RTT Annex is also recommended.

12.4 Relations to emergency service specifications

ETSI TC EMTEL has published standards about Total Conversation access to emergency services, and also general standards about access to IP-based emergency services where RTT is included. These documents may benefit from analysis and modifications for a clearer specification of how multiparty RTT calling would be implemented in emergency service use cases.

One case where there is an apparent need for multiparty RTT support is for the case when agencies within the emergency service want to establish multiparty calls for expert assistance or for assisted call transfer. Another case is for inclusion of language translation by interpreters or relay services in the call. This is described in clause 11.5 of the present document. Target documents would include ETSI TS 101 470 [i.2], ETSI TR 103 201 [i.3] and ETSI TS 103 479 [i.5] or as EMTEL experts find appropriate. In most cases, when the documents refer to IETF RFC 4103 [i.23] for RTT initiation and transport, the modifications would add that IETF RFC 4103 [i.23] has an update for multiparty calling in IETF RFC 9071 [i.30]. It may also be appropriate to refer to EN 301 549 [i.13] for RTT requirements.

In the Pan-European Mobile Emergency Application (PEMEA) system ETSI TS 103 478 [i.4], the RTT specification is in progress in TC EMTEL as draft ETSI TS 103 871 [i.36]. It includes functionality for multiparty calling.

13 Proposed information in EN 301 549

13.1 The need to update EN 301 549 to better address RTT

There will be a need for some modification of the existing requirements in clauses 6 and 13 of EN 301 549 [i.13] as well as the need for new requirements to be added. Some changes will be needed to more fully address important RTT features, such as the need to be able to modify previously entered text and to review incoming text that may have not been read immediately it was received. Other requirements will be needed that are important in the case of two-party RTT but are of greatest significance in multiparty RTT and Total Conversation scenarios.

Changes to existing requirements are likely to be largely confined to things like the clarification of some of the associated notes and the removal of some notes that will now be directly addressed by new requirements in clause 6.2.

NOTE 1: All references to clauses in clause 13 of the present document refer to clauses in EN 301 549 [i.13] version 3.2.1 unless they are explicitly preceded by the words "in the present document".

NOTE 2: All references to documents in the citations in clause 13 of the present document refer to reference documents either in EN 301 549 [i.13] version 3.2.1 or proposed to be included in EN 301 549 [i.13].

13.2 Changes to informative references in clause 2.2

The following new informative references are proposed to be inserted in clause 2.2.

"[i.rfc8825] *IETF RFC8825 (2021): "Overview: Real-Time Protocols for Browser-Based Applications"*

[i.rfc8865] *IETF RFC 8865 (2021): "T.140 Real-Time Text Conversation over WebRTC Data Channels"*

[i.rfc9071] *IETF RFC 9071 (2021): "RTP-Mixer Formatting of Multiparty Real-Time Text"*

[i.iso10646] *ISO/IEC 10646:2020: "Information technology - Universal coded character set (UCS)"*

[i.ts103479] *ETSI TS 103 479: "Emergency Communications (EMTEL); Core elements for network independent access to emergency services".*

Reference *[i.12]* has an error and is proposed to be corrected to:

"[i.12] *ETSI TS 124 229: "Universal Mobile Telecommunications System (UMTS); 5G; IP multimedia call control protocol based on Session Initiation Protocol (SIP)and Session Description Protocol (SDP); Stage 3 (3GPP TS 24.229)."*

13.3 Changes to definitions of terms in clause 3.1

1) A new term "continuous real-time conversation" is proposed to be added to replace the tentatively misleading "two-way communication" used in clause 6.2.1.1:

*"**continuous real-time conversation:** type of organization in conversation and discourse where one participant's contribution is made available to other participants while it is being made"*

2) In the definition of the term "Real-Time Text (RTT), note 1, it is proposed to change the allowed END-TO-END delay from 500 ms to 1 s, in order to align with the requirements for good conversational text (T2) from Recommendation ITU-T F.700 [i.33], and with the allowed 500 ms delay BEFORE TRANSMISSION responsiveness clause 6.2.4:

"NOTE 1: Users will perceive communication as continuous if the delay between text being created by the sender and received by the recipient is less than 1 s. However, the actual delay will be dependent on the communication network."

13.4 Changes to existing requirements in clause 6

13.4.1 Continuous real-time conversation

Clauses 6, 6.1 and 6.2.1.1 on RTT Communication contain the term "two-way communication." Even if this term "two-way" means both sending and receiving for all parties in the session, it may inadvertently give the impression that it is about communication between just two parties. That should be avoided because the requirements are valid also for multiparty calls. A more general and not misleading term is "continuous real-time conversation," which is proposed to be added to the terms and replace "two-way" in clauses 6, 6.1 and 6.2.1.1 and 6.2.1.2. In the header of clause 6 and in clause 6.1, it is sufficient to replace "two-way" with "conversational real-time" in all occurrences.

It is proposed that the words "in order to provide good audio quality, that" be deleted from the clause 6.1 requirement and that following note is proposed to be added to it:

"NOTE 1: The frequency range provides sufficient audio quality for reliable voice perception by most call participants."

Clause 6.2.1.1 also does not need to be tied to voice communication and is proposed to be changed to:

"Where ICT provides functionality that allows continuous real-time conversation, the ICT shall provide functionality that allows RTT communication."

The word "voice" needs to be retained in clause 6.2.1.2.

13.4.2 RTT communication

Clause 6.2.1.1 introduces RTT, but does not define it. The reader needs to review the definition of Real-Time Text to find what it means. To reduce the risk that the requirement to support multiparty calling is missed, the following changes are proposed. The main clause to read:

"Where ICT provides functionality that allows continuous real-time conversation, the ICT shall provide functionality that allows RTT communication, except where this would require design changes to add input or output hardware to the ICT."

The following note to be added:

"NOTE 4: The definition of Real-Time Text implies support of multiparty RTT communication."

13.4.3 Concurrent voice and text

Clause 6.2.1.2 on "Concurrent voice and text" contains a Note 1 which is not very clearly expressed and will benefit from a rephrasing. A rewording of note 3 is proposed to make it easier to read. In note 4, it is proposed that the word "field" be deleted. In note 5, "sold" and "product" are proposed to be replaced by "provided" and "ICT system". The changed and added NOTES 1,3 and 5 of 6.2.1.2 are proposed to be worded as below while the other notes are proposed to be kept as-is except that all occurrences of "many-party" should be replaced by "multiparty":

"NOTE 1: With multiparty communication, such as a conference system, conference floor control can be used to avoid confusion caused by multiple participants contributing at the same time.

NOTE 3: With a multiparty conference system that has chat as one of its features; RTT (like voice) would typically be separate from the chat so that RTT use does not interfere with chat. This separation allows participants to message in the chat field while a person contributes in real-time by RTT. This corresponds to the way participants message using chat while participants talk with voice. RTT users would then use RTT for presenting and use the chat feature to message while others are presenting by voice or RTT.

NOTE 5: Where both server-side software and local hardware and software are required to provide voice communication, where neither party can support voice communication without the other and are provided as a unit for the voice communication function, the local and server-side components are considered a single ICT system.

13.4.4 Distinguishable display

The text and note in clause 6.2.2.1 could be interpreted as if it would be acceptable to merge characters from different sources which would lead to unreadability and is against the title, the requirement of the clause and the basic characteristics of RTT. This is even more apparent for the multiparty case. The text and the note should be reworded to instead lead to provision of good usability for braille users with the principles of the clause maintained.

The title should be changed from "Visually distinguishable display" to "Distinguishable display" to also include tactile display. The new wording proposal for clause 6.2.2.1 and its notes is:

"Where ICT has RTT presentation capabilities, displayed received text from different sources and sent text shall, by default, be separated with their sources indicated and differentiated.

NOTE 1: The ability of the user to choose between different layouts of sent text and the text from the different sources, still fulfilling the requirement in this clause, allows users to display RTT in a form that works best for them.

NOTE 2: "Separated" here means presented in chunks as determined by common ICT conventions and language/locale readability expectations. Such chunks are usually either a completed response or, if the completed response is very long, chunks may be subdivided into a reasonably understandable natural language clause, phrase, or sentence."

13.4.5 Programmatically determinable send and receive origination

Clause 6.2.2.2 on "Programmatically determinable send and receive direction" requires minor modification to clearly be valid for multiparty calling. The title is proposed to be changed to " Programmatically determinable send and receive origination". The text and note of clause 6.2.2.2 are proposed to be changed to:

"Where ICT has RTT send and receive capabilities, the origin of text shall be programmatically determinable, unless the RTT is implemented as closed functionality.

NOTE: This enables screen readers to distinguish between incoming text from different sources and outgoing text when used with RTT functionality."

13.4.6 User identification

Clause 6.2.2.3 on "Speaker identification" already establishes the principle that participants using RTT should be identified if those using voice are identified. In voice communication some level of identification can often be derived from the characteristics of the spoken input. There is no realistic equivalent to such implicit identification of RTT users, so explicit identification of them becomes far more important. Since the ability of RTT users to identify voices may be limited, the identification of speakers and other active participants becomes more important in calls with RTT included.

The importance of explicit identification increases as the number of participants increase. This is true for voice users but is even more so for text users. This all points to the need to make identification of RTT users an absolute requirement and not one that is dependent on whether identification of voice uses exists or not.

In multiparty communication the presence of identifiers for all participants, whether they are using voice, RTT, or (in Total Conversation services) video to communicate, allows a comprehensive list of conference participants to be created, independent of their chosen communication modality. Such a listing greatly reduces the risk that some participants will be less "visible" to other participants who are not using the same modality.

It is proposed that the sentence in clause 6.2.2.3 is made independent of voice and changed to read:

"Where ICT has RTT capabilities, the ICT shall provide identification of the active RTT participant."

13.4.7 Visual indication of Audio with RTT

Clause 6.2.2.4 has a requirement to show audio activity in a call including RTT. The only change proposed in this clause is to change "two-party" to "conversational real-time" in the first line of the clause for the reasons indicated in clause 13.4.1 in the present document.

13.4.8 Additional clauses about RTT presentation

There are a number of valid additional requirements which are recommended to be placed on RTT presentation both for multiparty and two-party use. They are collected from other documents about RTT, and are proposed to be inserted here numbered 6.2.2.5 to 6.2.2.8:

"6.2.2.5 Use of RTT with Assistive Technology

Where ICT has assistive technology and includes RTT presentation capabilities, the ICT shall make all informational, navigational, and operational capabilities of the RTT interface programmatically determinable and executable by assistive technologies.

6.2.2.6 Presentation means for RTT

Where ICT has RTT presentation capabilities, the ICT shall provide RTT presentation in all ways available for general text presentation by the ICT.

6.2.2.7 Reviewing of received RTT

Where ICT has RTT presentation capabilities, the ICT shall provide the ability to review previously received RTT input for the current or latest call.

NOTE 1: When the presentation is in a position for presenting earlier communicated text, the local user usually prefers that the reading position is only changed by local user action and not by arrival of new text, while when the presentation is positioned to present latest text, then arrival or transmission of new text is usually preferred to be automatically presented.

NOTE 2: Ideally, the ability to review the last call will last for at least 24 hours.

6.2.2.8 Presentation of relative time order of text

Where ICT has RTT presentation capabilities, the presenting ICT shall provide the ability to present text with the approximate relative order in time when text from the different sources (including the local sending user) appeared.

NOTE: The requirement is for an approximate time order. When texts from different parties have been received simultaneously, showing an exact time relation between text entry from different parties is unnecessary and not suitable to maintain readability of the text in most presentation layout alternatives.

6.2.2.9 Presentation of related text from different sources

Where ICT has RTT presentation capabilities, the ICT shall present text from each source separately and divided in groups at places in the text where the division is not likely to disturb reading."

13.4.9 Interoperability

The interoperability clause 6.2.3 specifies how interoperability can be achieved between ICT with RTT functionality. It lists the main session and transport level standards for RTT in the most common call control protocol environments. After publication of version 3.2.1 of EN 301 549 [i.13], there has been progress in standardisation of RTT including multiparty aspects, for the two most common communication environments: SIP and WebRTC. This makes it possible to add details to the list of interoperability methods in clause 6.2.3.

Item a) in the list in clause 6.2.3 covers PSTN text telephony, but no implementation of this provides full RTT functionality. The different kinds of implemented text telephones collected under item a) show different limitations compared to RTT and no one supports multiparty calling natively. Only a few countries still support PSTN text telephones, and more countries are expected to drop legacy support requirements. Because PSTN text telephony is a legacy system that is not RTT, it does not belong in a clause that is describing RTT interoperability mechanisms.

It is proposed to:

- introduce a new heading "6.2.3 Interoperability and interworking";
- retitle the original clause 6.2.3 to become "6.2.3.1 RTT interoperability";
- modify the introduction to read:

  "Where ICT has RTT functionality (as required by clause 6.2.1.1) it shall support the applicable RTT interoperability mechanisms described below:";

- item a) in the original list to be deleted;
- the original item b) to be split in two cases a) and b), and a new item c) inserted before the original item c) and d), moved to be items d) and e) followed by notes and example;
- "a or b" changed to "a or b or c" in the original option c) (now option d));
- "option c" in the Example will be changed to "option d";
- The new items a), b), and c) are proposed to read:

  "a) ICT interoperating with other ICT using VOIP with Session Initiation Protocol (SIP) and using RTT that conforms to IETF RFC 4103 [i.13] including the updates for multiparty use in IETF RFC 9071 [i.rfc9071];

  b) ICT interoperating with other ICT using the IP Multimedia Sub-System (IMS) to implement VOIP with RTT using the set of protocols specified in ETSI TS 126 114 [i.10], ETSI TS 122 173 [i.11] and ETSI TS 124 229 [i.12] describing how IETF RFC 4103 [i.13] would apply updated by IETF RFC 9071 [i.rfc9071];

  c) ICT interoperating with other ICT using WebRTC [i.rfc8825] technology to implement VoIP using IETF RFC 8865 [i.rfc8865] to implement RTT functionality using web technologies."

- add a new clause "6.2.3.2 Legacy fall-back interworking with text telephony" to address those cases where there is a public or corporate policy to provide some form of interworking between PSTN text telephony and RTT. This proposed text for this clause is:

  "Where there is a public or corporate policy for ICT with RTT functionality to interwork with ICT supporting legacy PSTN text telephony, the ICT shall use Recommendation ITU-T V.18 [i.23] or an applicable annex thereof for text telephony signals at the PSTN interface for the case that there is a requirement to support that type of text telephone in the location of the PSTN user connection.

  NOTE 1: The functionality of interworking between PSTN text telephony and RTT cannot fulfil all RTT requirements in the present document.

  NOTE 2: No PSTN text telephone standard can properly support multiparty RTT presentation, and even where they are enabled to participate in multiparty RTT calls, PSTN textphones can only properly present text from one call participant at a time."

13.4.10 RTT responsiveness

The notes in clause 6.2.4 about "RTT responsiveness", may be a good place to add extra information about the benefits of RTT to motivate implementers to provide good RTT functionality. Note 1 is proposed to be extended for the case of composed characters. An added Note 4 is proposed to be added to clause 6.2.4:

"NOTE 1: For character-by-character input, the "smallest reliably composed unit of text entry" would be a character even if it is composed by multiple keystrokes. For word prediction it would be a word. For some voice recognition systems - the text may not exit the recognition software until an entire word (or phrase) has been spoken. In this case, the smallest reliably composed unit of text entry available to the ICT would be the word (or phrase).

......

NOTE 4: During emergency service applications, it is especially critical to send the smallest reliably composed unit of text entry within 500 ms, regardless of any user setting or preferences."

13.4.11 Further requirements on RTT

A set of other requirements on RTT have been derived from other standards. They are proposed to be inserted as clauses 6.2.5 to 6.2.11:

"6.2.5 Revising RTT

Where ICT utilises RTT input, the RTT user interface shall provide the participant with functionality to provide new input and to revise text by erasure from the end, without being limited by any New Line or other delimiter in the earlier text.

NOTE: "Functionality to revise text" applies to manual correction and automatic correction mechanisms such as automatic spelling correction.

6.2.6 Speed of RTT

Where ICT has RTT handling capabilities, the ICT shall be able to handle input from individual RTT users at a rate of at least 30 characters per second and reception and presentation of at least 30 characters per second from each actively sending RTT user.

NOTE: In the rare scenario where many users are sending RTT text simultaneously, ICT may adjust the practical limit for the character refresh rate.

6.2.7 Character representation

Where ICT has RTT input and/or RTT presentation capabilities, the ICT shall be capable of handling the characters of ISO/IEC 10646 [i.iso10646] or at least a subset thereof, including any supported emojis, the replacement character used for indication of unsupported characters, and the writing direction(s) matching the characters supported by the ICT device or applications in use.

6.2.8 RTT input methods

When ICT has RTT input capabilities, the ICT shall provide RTT input methods in all ways available for general text input by the ICT.

6.2.9 Media support

When ICT has RTT support, it shall be possible to include RTT from the beginning of the call (for both incoming and outgoing calls) and to add RTT during the call in any call leg.

6.2.10 Maintaining existing telephony functionality

All standard functions of calls with audio (e.g. mute audio, volume, touchtone buttons) shall also function when RTT is included, except where functions are disabled on emergency service calls due to safety concerns or regulations.

NOTE 1: For example, with multiparty RTT, this could include call hold, merging, swapping, and similar functionality.

NOTE 2: Clause 13.3.2 covers "Multiparty call support during emergency service calls."

13.5 Identification of actively speaking users

Identifying which users are actively contributing to the communication is a separate concept to merely identifying who is present in the communication (even though displaying who is actively contributing may sometimes make use of the same identity presentation elements). Identifying persons actively contributing is very valuable when all participants are communicating in the same modality, but it becomes even more essential when some of the contributors are using a modality that some other participants may not be continuously monitoring. Care will be needed to see what can be required in this area. This topic is addressed for video in clauses 6.5.5 and 6.5.6. These clauses should be reviewed to see if a solution that covers the needs for RTT, audio, video and text chat in one clause can be found.

The programmatic determinability would need to be enhanced from just determining the direction to determining the source of all characters of text so that text can be collected into suitable readable chunks from the same source and presented also in tactile and audible modes.

A proposed solution is to insert the following clause as 6.7:

"6.7 Activity indication

Where ICT provides real-time voice-based communication, the ICT shall provide accessible indications of activities in the supported media.

"NOTE 1 It is relevant to maintain the indication for activity in RTT until the user has positioned the reading position to the new text, while for voice and video, it is relevant to maintain the indication just as long as the activity is ongoing.

NOTE 2: The requirement in this clause can be deduced from the general accessibility requirements in the present document, but is expressed here as well.

NOTE 3: A suitable place for the indication could be in a participants list in a visual user interface, but also in dedicated positions in a tactile user interface."

13.6 Identifying who wishes to communicate next

A "raise hand" feature is very commonly available in conferencing systems to allow users to express a desire to contribute. Such a feature becomes even more important when participants may be contributing using different modalities. No one should be disadvantaged in expressing their desire to contribute because of their chosen communication modality. This suggests the need for a common system for all users, regardless of how they prefer to contribute.

A new clause 6.8 is proposed for this purpose:

"6.8 Conference floor control

Where ICT provides conference functionality for continuous real-time conversation, any conference floor control functionality provided by the ICT should allow a participant to indicate that they want to contribute, regardless of which media the participant intends to use.

> *NOTE: A hand-raising tool is a suitable feature to support coordination of speakers, signers, or text contributors using different media for their conference contributions."*

13.7 Awareness versus distraction

The variety of things that potentially need to be signalled to users to ensure that everyone is able to effectively participate, including everything referred to in the previous clauses, could potentially be overwhelming to some users who suffer from attentional issues. Some users might find themselves getting too focused on the notifications and find that it is distracting them from following the content of the discussions. Others may simply get overwhelmed by the frequency and variety of status information being directed at them and have a total attentional overload and be unable to continue participating.

There is no simple solution that will provide every user exactly what they want to be aware of and nothing more in the exact way that meets their individual needs. The only likely resolution to this dilemma will be a high degree of personalization to individual users. But a further dilemma is that the wider the scope of personalisation that is made available to users the more scope there is for having the user interface to personalize the user experience being a significant complicating factor.

As EN 301 549 [i.13] will never contain complete detailed user interface solutions, it is as yet unclear to what extent it will be possible to alert and guide those who provide the user interfaces to pay attention to trying to find the right balance between ultimate configurability and simplicity.

13.8 Programmatic determinability of the new notifications

Most of the examples in the present document are expressed in terms of providing text content and notifications in terms of their visual appearance on screens. Although this should be generally covered by other requirements in EN 301 549 [i.13], the availability of this information to all users, including Deaf-Blind users, should be implicit in all the new requirements added to the standard. This might sometimes be emphasised by strategically placed notes that remind developers of the obligation for the content and notifications to be available in an accessible way to all users.

13.9 Changes to requirements on services in clause 13

13.9.1 Access to relay services

In clause 13.2 about access to relay services, the expression "not be prevented" should be replaced with "be supported" because invoking a relay service in a call requires support of specific actions by the ICT, especially when the invocation is made seamlessly so that the call provides functional equivalence with calling between users of the same modality. The actions may for example include connection or the parties and the relay service in a multiparty call with media distribution specifically designed for the type of relay service and its mode of operation.

13.9.2 Access to emergency services

The single clause 13.3 about access to emergency services is not sufficient to describe the requirements for accessible access to emergency services. The current clause 13.3 is proposed to form the first clause 13.3.1 "Media support for emergency services". In this clause, "two-way" is proposed to be replaced with "conversational real-time", "not be prevented" with "be supported" with the same reasoning as in clause 13.9.1 of the present document, and "with each media" inserted before "individually" for clarification.

This new note is proposed to be added to clause 13.3.1:

"NOTE 3: An emergency service access standard supporting RTT is ETSI TS 103 479 [i.ts103479]."

The following two new clauses are also proposed to be inserted after clause 13.3.1 to clarify the most basic handling of accessible emergency calls:

"13.3.2 Multiparty call support during emergency service calls

Multiparty calling shall be supported during emergency calls when the initiative to add parties to the call comes from the emergency service, including calls using audio, RTT, or video, either individually or in any combination.

NOTE 1: Communication with emergency services commonly implies interacting in multiparty calls with the active media in the call (audio, video and RTT) and is therefore successfully served only by ICT supporting multiparty calling in the media used.

NOTE 2: Emergency access standards require some supplementary services, such as hold, conference and transfer, to be disabled from activation from the user device during emergency calls and callback.

NOTE 3: Emergency services prefer to get audio from the user terminal even when RTT is the main medium. Disabling muting of the microphone and audio transmission in user terminals during emergency calls and callback meets this need.

13.3.3 Callback during emergency calls

Callback from the emergency services to ICT operated by a user in an emergency shall be supported with the same or greater combination of conversational media (voice, video, RTT) as the original call to emergency services.

EXAMPLE 1: An emergency services callback for a Voice+RTT call would be expected to support both Voice and RTT.

EXAMPLE 2: An emergency service callback for a Voice+RTT+Video call would be expected to support Voice, Video, and RTT."

Annex A (informative): EN 301 549 extracts with proposed changes

A.1 Introduction

The present annex contains extracts from EN 301 549 [i.13] V3.2.1 with the proposed changes and additions described in clause 13 of the present document applied. Ellipsis "…" replace parts of clauses which are not included because they are not affected by the change proposals.

Text proposed to be added is made in red and italics. Text proposed to remain unchanged is black in its original font and style. Text proposed to be deleted is marked by a double overstrike.

To fully understand the implications of the proposed new and modified requirements in the present annex, it may be helpful to read the following proposals whilst having access to version 3.2.1 of EN 301 549 [i.13].

NOTE 1: All references to clauses in Annex A.2 of the present document refer to clauses in EN 301 549 [i.13] version 3.2.1.

NOTE 2: All references to documents in Annex A.2 of the present document refer to reference documents either in EN 301 549 [i.13] version 3.2.1 or proposed to be included in EN 301 549 [i.13].

A.2 The EN 301 549 revisions and additions

----------NO MODIFICATIONS PROPOSED IN CLAUSES PRIOR TO CLAUSE 2 ----------

2 References

----------NO MODIFICATIONS PROPOSED IN CLAUSE 2.1 ----------

2.2 Informative references

. . .

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

…

[i.rfc8825] *IETF RFC8825 (2021): "Overview: Real-Time Protocols for Browser-Based Applications"*

[i.rfc8865] *IETF RFC 8865 (2021): "T.140 Real-Time Text Conversation over WebRTC Data Channels"*

[i.rfc9071] *IETF RFC 9071 (2021): "RTP-Mixer Formatting of Multiparty Real-Time Text"*

[i.iso10646] *ISO/IEC 10646:2020: "Information technology - Universal coded character set (UCS)"*

[i.ts103479] *ETSI TS 103 479: "Emergency Communications (EMTEL); Core elements for network independent access to emergency services".*

. . .

[i.12] ETSI TS 1~~3~~*2*4 229: "Universal Mobile Telecommunications System (UMTS);~~LTE; Internet Protocol (IP) multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); User Equipment (UE) conformance specification~~ *5G;IP multimedia call control protocol based on Session Initiation Protocol (SIP)and Session Description Protocol (SDP);Stage 3 (3GPP TS 2*~~3~~*4.229)".*

3 Definition of terms, symbols and abbreviations

3.1 Terms

…

continuous real-time conversation: *type of organization in conversation and discourse where one participant's contribution is made available to other participants while it is being made*

…

Real-Time Text (RTT): form of a 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

NOTE 1: Users will perceive communication as continuous if the delay between text being created by the sender and received by the recipient is less than ~~500 ms~~*1 s*. However, the actual delay will be dependent on the communication network.

NOTE 2: The creation of text will differ between systems where text is entered on a word-by-word basis (e.g. speech-to-text and predictive-text based systems) and systems where each character is separately generated (e.g. typing on a physical keyboard).

. . .

----------NO MODIFICATIONS PROPOSED IN CLAUSES 3.2 - 5 ----------

6 ICT with conversational real-time voice communication

6.1 Audio bandwidth for speech

Where ICT provides *conversational real-time* ~~two-way~~ voice communication~~, in order to provide good audio quality, that~~ ICT shall be able to encode and decode *conversational real-time* ~~two-way~~ voice communication with a frequency range with an upper limit of at least 7 000 Hz.

NOTE 1: The frequency range provides sufficient audio quality for reliable voice perception by most call participants.

NOTE ~~1~~*2*: For the purposes of interoperability, support of Recommendation ITU-T G.722 [i.21] is widely used.

NOTE ~~2~~*3*: Where codec negotiation is implemented, other standardized codecs such as Recommendation ITU-T G.722.2 [i.22] are sometimes used so as to avoid transcoding.

6.2 Real-Time Text (RTT) functionality

6.2.1 RTT provision

6.2.1.1 RTT communication

Where ICT provides functionality that ~~provides a means for two-way voice communication~~ *allows continuous real-time conversation*, the ICT shall ~~provide a means for two-way RTT communication~~ *provide functionality that allows RTT communication*, except where this would require design changes to add input or output hardware to the ICT.

NOTE 1: This requirement includes those products which do not have physical display or text entry capabilities but have the capability to connect to devices that do have such capabilities. It also includes intermediate ICT between the endpoints of the communication.

NOTE 2: There is no requirement to add: a hardware display, a hardware keyboard, or hardware to support the ability to connect to a display or keyboard, wired or wirelessly, if this hardware would not normally be provided.

NOTE 3: For the purposes of interoperability, support of Recommendation ITU-T T.140 [i.36] is widely used.

NOTE 4: The definition of Real-Time Text implies support of multiparty RTT communication.

6.2.1.2 Concurrent voice and text

Where ICT provides a means for ~~two-way~~*continuous real-time* voice ~~communication~~ *conversation* and for users to communicate by RTT, it shall allow concurrent voice and text through a single user connection.

NOTE 1: With ~~many-party~~ *multiparty* communication, ~~as in~~*such as* a conference system, ~~it is allowed (but not required or necessarily recommended) that RTT be handled in a single display field and that "turn-taking" be necessary~~ *conference floor control can be used* to avoid confusion ~~(in the same way that turn-taking is required for those presenting/talking with voice)~~*caused by multiple participants contributing at the same time*.

NOTE 2: With multiparty communication, best practice is for hand-raising for voice users and RTT users to be handled in the same way, so that voice and RTT users are in the same queue.

~~NOTE 3: With a multiparty conference system that has chat as one of its features - the RTT (like the voice) would typically be separate from the chat so that RTT use does not interfere with chat (i.e. participants can be messaging in the chat field while the person is presenting/talking with RTT - in the same manner that participants message using the chat feature while participants are talking with voice). RTT users would then use RTT for presenting and use the Chat feature to message while others are presenting (via Voice or RTT).~~

NOTE 3: With a multiparty conference system that has chat as one of its features; RTT (like voice) would typically be separate from the chat so that RTT use does not interfere with chat. This separation allows participants to message in the chat field while a person contributes in real-time by RTT. This corresponds to the way participants message using chat while participants talk with voice. RTT users would then use RTT for presenting and use the chat feature to message while others are presenting by voice or RTT.

NOTE 4: The availability of voice and RTT running concurrently (and separately from chat) can also allow the RTT ~~field~~ to support text captioning when someone is speaking (and it is therefore not being used for RTT since it is not the RTT user's turn to speak).

NOTE 5: Where both server-side software and local hardware and software are required to provide voice communication, where neither part can support voice communication without the other and are ~~sold~~*provided* as a unit for the voice communication function, the local and server-side components are considered a single ~~product~~ *ICT system*.

6.2.2 Display of RTT

6.2.2.1 Distinguishable display

Where ICT has RTT ~~send and receive~~*presentation* capabilities, displayed ~~sent~~ *received* text ~~shall be visually differentiated from, and separated from, received text~~ from different sources and sent text *shall, by default, be separated with their sources indicated and differentiated*.

NOTE *1*: The ability of the user to choose between ~~having the send and receive text be displayed in-line or separately, and with options to select, allows users to display RTT in a form that works best for them. This would allow Braille users to use a single field and take turns and have text appear in the sequential way that they may need or prefer~~ *different layouts of sent text and the text from the different sources, still fulfilling the requirement in this clause, allows users to display RTT in a form that works best for them*.

NOTE 2: "Separated" here means presented in chunks as determined by common ICT conventions and language/locale readability expectations. Such chunks are usually either a completed response or, if the completed response is very long, chunks may be subdivided into a reasonably understandable natural language clause, phrase, or sentence.

6.2.2.2 Programmatically determinable send and receive ~~direction~~ *origination*

Where ICT has RTT send and receive capabilities, ~~the send/receive direction of transmitted/received text~~ *the origin of text* shall be programmatically determinable, unless the RTT is implemented as closed functionality.

NOTE: This enables screen readers to distinguish between incoming text *from different sources* and outgoing text when used with RTT functionality.

6.2.2.3 Speaker identification

Where ICT has RTT ~~send and receive~~ capabilities~~, and provides speaker identification for voice,~~ the ICT shall provide ~~speaker~~ identification ~~for~~ *of the active* RTT *participant*.

NOTE: This is necessary to enable both voice and RTT participants to know who is currently communicating, whether it be in RTT or voice.

6.2.2.4 Visual indicator of Audio with RTT

Where ICT provides ~~two-way voice~~ *conversational real-time voice* communication, and has RTT capabilities, the ICT shall provide a real-time visual indicator of audio activity on the display.

NOTE 1: The visual indicator may be a simple character position on the display that flickers on and off to reflect audio activity, or presentation of the information in another way that can be both visible to sighted users and passed on to deaf-blind users who are using a braille display.

NOTE 2: Without this indication a person who lacks the ability to hear does not know when someone is talking.

6.2.2.5 Use of RTT with Assistive Technology

Where ICT has assistive technology and includes RTT presentation capabilities, the ICT shall make all informational, navigational, and operational capabilities of the RTT interface programmatically determinable and executable by assistive technologies.

6.2.2.6 Presentation means for RTT

Where ICT has RTT presentation capabilities, the ICT shall provide RTT presentation in all ways available for general text presentation by the ICT.

6.2.2.7 Reviewing of received RTT

Where ICT has RTT presentation capabilities, the ICT shall provide the ability review previously received RTT input for the current or latest call.

NOTE 1: When the presentation is in a position for presenting earlier communicated text, the local user usually prefers that the reading position is only changed by local user action and not by arrival of new text, while when the presentation is positioned to present latest text, then arrival or transmission of new text is usually preferred to be automatically presented.

NOTE 2: Ideally, the ability to review the last call will last for at least 24 hours.

6.2.2.8 Presentation of relative time order of text

Where ICT has RTT presentation capabilities, the presenting ICT shall provide the ability to present text with the approximate relative order in time when text from the different sources (including the local sending user) appeared.

NOTE: The requirement is for an approximate time order. Showing exact time relation between text entries from different parties would not be needed and not suitable in most possible presentation layout alternatives with maintained readability of the texts when texts from different parties have been received simultaneously.

6.2.2.9 Presentation of related text from different sources

Where ICT has RTT presentation capabilities, the ICT shall present text from each source separately and divided in groups at places in the text where the division is not likely to disturb reading.

6.2.3 Interoperability *and interworking*

6.2.3.1 RTT interoperability

Where ICT *has* ~~with~~ RTT functionality ~~interoperates with other ICT with RTT functionality~~ (as required by clause 6.2.1.1)~~ they~~ *it* shall support the applicable RTT interoperability mechanisms described below:

~~a) ICT interoperating with other ICT directly connected to the Public Switched Telephone Network (PSTN), using Recommendation ITU-T V.18 [i.23] or any of its annexes for text telephony signals at the PSTN interface;~~

~~b)~~*a)* ICT interoperating with other ICT using VOIP with Session Initiation Protocol (SIP) and using RTT that conforms to IETF RFC 4103 [i.13] *including the updates for multiparty use in IETF RFC 9071[i.rfc9071]*~~. For ICT interoperating with other ICT using the IP Multimedia Sub-System (IMS) to implement VOIP, the set of protocols specified in ETSI TS 126 114 [i.10], ETSI TS 122 173 [i.11] and ETSI TS 134 229 [i.12] describe how IETF RFC 4103 [i.13] would apply~~;

b) ICT interoperating with other ICT using the IP Multimedia Sub-System (IMS) to implement VOIP with RTT using the set of protocols specified in ETSI TS 126 114 [i.10], ETSI TS 122 173 [i.11] and ETSI TS 124 229 [i.12] describing how IETF RFC 4103 [i.13] would apply updated by IETF RFC 9071 [i.rfc9071];

c) ICT interoperating with other ICT using WebRTC [i.rfc8825] technology to implement VoIP using IETF RFC 8865 [i.rfc8865] to implement RTT functionality using web technologies.

~~c)~~*d)* ICT interoperating with other ICT using technologies other than a or b *or c*, above, using a relevant and applicable common specification for RTT exchange that is published and available for the environments in which they will be operating. This common specification shall include a method for indicating loss or corruption of characters.

~~d)~~*e)* ICT interoperating with other ICT using a standard for RTT that has been introduced for use in any of the above environments, and is supported by all of the other active ICT that support voice and RTT in that environment.

NOTE 1: In practice, new standards are introduced as an alternative codec/protocol that is supported alongside the existing common standard and used when all end-to-end components support it while technology development, combined with other reasons including societal development and cost efficiency, may make others become obsolete.

NOTE 2: Where multiple technologies are used to provide voice communication, multiple interoperability mechanisms may be needed to ensure that all users are able to use RTT.

EXAMPLE: A conferencing system that supports voice communication through an internet connection might provide RTT over an internet connection using a proprietary RTT method (option ~~c~~*d*). However, regardless of whether the RTT method is proprietary or non-proprietary, if the conferencing system also offers telephony communication it will also need to support option a or b to ensure that RTT is supported over the telephony connection.

6.2.3.2 Legacy fall-back interworking with text telephony

Where there is a public or corporate policy for ICT with RTT functionality to interwork with ICT supporting legacy PSTN text telephony, the ICT shall use Recommendation ITU-T V.18 [i.23] or an applicable annex thereof for text telephony signals at the PSTN interface for the case that there is a requirement to support that type of text telephone in the location of the PSTN user connection.

NOTE 1: The functionality of interworking between PSTN text telephony and RTT cannot fulfil all RTT requirements in the present document.

NOTE 2: No PSTN text telephone standard can properly support multiparty RTT presentation, and even where they are enabled to participate in multiparty RTT calls, PSTN textphones can only properly present text from one call participant at a time.

6.2.4 RTT responsiveness

Where ICT utilises RTT input, that RTT input shall be transmitted to the ICT network or platform on which the ICT runs within 500 ms of the time that the smallest reliably composed unit of text entry is available to the ICT for transmission. Delays due to platform or network performance shall not be included in the 500 ms limit.

NOTE 1: For character-by-character input, the "smallest reliably composed unit of text entry" would be a character *even if it is composed by multiple keystrokes*. For word prediction it would be a word. For some voice recognition systems - the text may not exit the recognition software until an entire word (or phrase) has been spoken. In this case, the smallest reliably composed unit of text entry available to the ICT would be the word (or phrase).

NOTE 2: The 500 ms limit allows buffering of characters for this period before transmission so character by character transmission is not required unless the characters are generated more slowly than 1 per 500 ms.

NOTE 3: A delay of 300 ms, or less, produces a better impression of flow to the user.

NOTE 4: During emergency service applications, it is especially critical to send the smallest reliably composed unit of text entry within 500 ms, regardless of user setting or preference.

6.2.5 Revising RTT

Where ICT utilises RTT input, the RTT user interface shall provide the participant with functionality to provide new input and to revise text by erasure from the end, without being limited by any New Line or other delimiter in the earlier text.

NOTE: "Functionality to revise text" applies to manual correction and automatic correction mechanisms such as automatic spelling correction.

6.2.6 Speed of RTT

Where ICT has RTT handling capabilities, the ICT shall be able to handle input from individual RTT users at a rate of at least 30 characters per second and reception and presentation of at least 30 characters per second from each actively sending RTT user.

NOTE: In the rare scenario where many users are sending RTT text simultaneously, ICT may adjust the practical limit for the character refresh rate.

6.2.7 Character representation

Where ICT has RTT input and/or RTT presentation capabilities, the ICT shall be capable of handling the characters of ISO/IEC 10646 [i.iso10646] or at least a subset thereof, including any supported emojis, the replacement character used for indication of unsupported characters, and the writing direction(s) matching the characters supported by the ICT device or applications in use.

6.2.8 RTT input methods

When ICT has RTT input capabilities, the ICT shall provide RTT input methods in all ways available for general text input by the ICT.

6.2.9 Media support

When ICT has RTT support, it shall be possible to include RTT from the beginning of the call (for both incoming and outgoing calls) and to add RTT during the call in any call leg.

6.2.10 Maintaining existing telephony functionality

All standard functions of calls (e.g. mute audio, volume, touchtone buttons) shall also function when RTT is included, except where functions are disabled in emergency services calls due to safety concerns or regulations.

NOTE 1: For example, with multiparty RTT, this could include call hold, merging, swapping, and similar functionality.

NOTE 2: Clause 13.3.2 covers "Multiparty call support during emergency service calls"

----------NO MODIFICATIONS PROPOSED IN CLAUSES 6.3 - 6.6 ----------

6.7 Activity indication

Where ICT provides real-time voice-based communication, the ICT shall provide accessible indications of activities in the supported media.

NOTE 1: It is relevant to maintain the indication for activity in RTT until the user has positioned the reading position to the new text, while for voice and video, it is relevant to maintain the indication just as long as the activity is ongoing.

NOTE 2: The requirement in this clause can be deduced from the general accessibility requirements in the present document, but is expressed here as well.

NOTE 3: A suitable place for the indication could be in a participants list in a visual user interface, but also in dedicated positions in a tactile user interface.

6.8 Conference floor control

Where ICT provides conference functionality for continuous real-time conversation, any conference floor control functionality provided by the ICT should allow a participant to indicate that they want to contribute, regardless of which media the participant intends to use.

NOTE: A hand-raising tool is a suitable feature to support coordination of speakers, signers, or text contributors using different media for their conference contributions.

----------NO MODIFICATIONS PROPOSED IN CLAUSES 7 - 12 ----------

13 ICT providing relay or emergency service access

----------NO MODIFICATIONS PROPOSED IN CLAUSE 13.1 ----------

13.2 Access to relay services

Where ICT systems support two-way communication, and the system is specified for use with relay services, access to those relay services shall ~~not be prevented~~ *be supported* for outgoing and incoming calls involving voice, RTT, or video, either individually or in combinations supported by both the relay service and the ICT system.

NOTE 1: The purpose of this requirement is to achieve functionally equivalent communication access by persons with disabilities.

NOTE 2: The system may be specified as needing to work with relay services by, for example: procurers, regulators, or product specifications.

13.3 Access to emergency services

13.3.1 Media support for emergency services

Where ICT systems support *conversational real-time* ~~two-way~~ communication, and the system is specified for use with emergency services, access to those emergency services shall *be supported* ~~not be prevented~~ for outgoing and incoming calls involving voice, RTT, or video, either *with each media* individually or in combinations supported by both the emergency service and the ICT system.

NOTE 1: The purpose of this requirement is to achieve functionally equivalent communication access to the emergency service by persons with disabilities.

NOTE 2: The system may be specified as needing to work with emergency services by, for example: procurers, regulators, or product specifications.

NOTE 3: An emergency service access standard supporting RTT is ETSI TS 103 479 [i.ts103479].

13.3.2 Multiparty call support during emergency service calls

Multiparty calling shall be supported during emergency calls when the initiative to add parties to the call comes from the emergency service, including calls using audio, RTT, or video, either individually or in any combination.

NOTE 1: Communication with emergency services commonly implies interacting in multiparty calls with the active media in the call (audio, video and RTT) and is therefore successfully served only by ICT supporting multiparty calling in the media used.

NOTE 2: Emergency access standards require some supplementary services, such as hold, conference and transfer, to be disabled from activation from the user device during emergency calls and callback.

NOTE 3: Emergency services prefer to get audio from the user terminal even when RTT is the main medium. Disabling muting of the microphone and audio transmission in user terminals during emergency calls and callback meets this need.

13.3.3 Callback during emergency calls

Callback from the emergency services to ICT operated by a user in an emergency shall be supported with the same or greater combination of conversational media (voice, video, RTT) as the original call to emergency services.

EXAMPLE 1: An emergency services callback for a Voice+RTT call would be expected to support both Voice and RTT.

EXAMPLE 2: An emergency service callback for a Voice+RTT+Video call would be expected to support Voice, Video, and RTT."

----------NO MODIFICATIONS PROPOSED BEYOND CLAUSE 13 ----------

Annex B (informative): Bibliography

- Text telephony for all (DUST). Walter Mellors, Scott Cadzow, Edward Fitzgerald, Gunnar Hellström, HFT'06, Human Factors in Telecommunications 2006.
- 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)".
- ETSI TS 122 101: "Universal Mobile Telecommunications System (UMTS); LTE; Service aspects; Service principles (3GPP TS 22.101)".
- ATIS-0700029 Real Time Text Mobile Device Behavior, ATIS 2017.

NOTE: Available at https://webstore.ansi.org/standards/atis/atis0700029.

- ATIS-0700030 Real Time Text End-to-End Service Description Specification, ATIS 2018.

NOTE: Available at https://webstore.ansi.org/standards/atis/atis0700030.

- ETSI ETR 333 (1977): "Text Telephony, User Requirements And Recommendations".

NOTE: Available at https://www.etsi.org/deliver/etsi_etr/300_399/333/01_60/etr_333e01p.pdf.

- ETSI TR 102 974 (V1.1.1): "Human Factors (HF); Telecommunications relay services" (Background information to ETSI ES 202 975).

NOTE: Available at https://www.etsi.org/deliver/etsi_tr/102900_102999/102974/01.01.01_60/tr_102974v010101p.pdf.

- ETSI EG 202 320 (V1.2.1): "Human Factors (HF); Duplex Universal Speech and Text (DUST) communications".

NOTE: Available at https://www.etsi.org/deliver/etsi_eg/202300_202399/202320/01.02.01_60/eg_202320v010201p.pdf.

- Recommendation ITU-T F.703: "Multimedia conversational services". ITU-T 2000.

NOTE: Available at https://www.itu.int/rec/T-REC-F.703-200011-I/en.

- IETF draft-hellstrom-textpreview-08 (2011): "Presentation of Text Conversation in real-time and en-bloc form".

NOTE: Available at https://datatracker.ietf.org/doc/html/draft-hellstrom-textpreview.

- W3C® Working Group Note: "RTC Accessibility User Requirements".

NOTE: Available at https://w3c.github.io/apa/raur/ as an Editor's draft version from August 17, 2021 to at some later time become officially available at https://www.w3.org/TR/raur/ when this link can replace the one above.

History

Document history		
V1.1.1	August 2022	Publication