ETSI TR 103 708 V0.0.8 (2022-01)

**RTT Multi-party conference calling Specification**

supporting RTT over multi-party telephony VoIP calling

<

**TECHNICAL REPORT**

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 - NAF 742 C

Association à but non lucratif enregistrée à la

Sous-préfecture de Grasse (06) N° 7803/88

***Important notice***

The present document can be downloaded from:
<http://www.etsi.org/standards-search>

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at [www.etsi.org/deliver](http://www.etsi.org/deliver).

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at <https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx>.

If you find errors in the present document, please send your comment to one of the following services:
<https://portal.etsi.org/People/CommiteeSupportStaff.aspx>

***Copyright Notification***

No part may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm except as authorized by written permission of ETSI.

The content of the PDF version shall not be modified without the written authorization of ETSI.

The copyright and the foregoing restriction extend to reproduction in all media.

© ETSI yyyy.

All rights reserved.

**DECT**TM, **PLUGTESTS**TM, **UMTS**TM and the ETSI logo are trademarks of ETSI registered for the benefit of its Members.
**3GPP**TM and **LTE**TM 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.

Contents

Intellectual Property Rights 6

Foreword 6

Modal verbs terminology 6

Executive summary 6

Introduction 6

1 Scope 6

2 References 7

2.1 Normative references 7

2.2 Informative references 7

3 Definition of terms, symbols and abbreviations 8

3.1 Terms 8

3.2 Symbols 9

3.3 Abbreviations 9

4 Real-time text (RTT) 9

4.1 What is RTT? 9

4.2 RTT is for conversation not messaging 9

5 Multiparty calls 9

6 User interface elements and use cases 10

6.1 Background 10

6.2 Types of media 10

6.2.1 Real-time text (RTT) 10

6.2.2 Audio 10

6.2.3 Video 10

6.2.4 Total Conversation 11

6.2.5 Messaging 11

6.2.6 Documents, drawings and stored or streamed media 11

6.3 Use of media and user interface elements 11

6.3.1 Potential user interface components 11

6.3.2 User generated media presentation 11

6.3.3 Active and waiting user activity indication 11

6.3.4 User control of presentation and input 12

6.4 Use cases 12

6.4.1 Use case variations 12

6.4.2 Call using RTT within a small group of Deaf persons 12

6.4.3 Deaf person calling emergency service and using RTT 12

6.4.4 Hard-of-hearing user talking with hearing friends 13

6.4.5 Deaf user participating in conference getting transcription support 13

6.4.6 Deaf-Blind user participating in remote meeting 13

6.4.7 Person in a critical situation making an emergency call by RTT 13

7 RTT text creation and presentation 14

7. 1 Text creation and transmission 14

7. 1.1 Variation in text creation means 14

7.1.2 Verification of produced text 14

7.1.3 Erasure 14

7.1.4 Input language and character set 14

7.1.5 Sending of emoji 15

7.1.6 Graphic rendition 15

7.1.7 Time from text entry to presentation 15

7.1.8 Support for user to send text when the situation allows 15

7.1.9 Rather loss than duplication in transport of text 15

7.2 RTT text presentation 16

7.2.1 RTT text presentation basics 16

7.2.2 Visual presentation variants 16

7.2.2.1 One column view 16

7.2.2.2 One column per contributing user 16

7.2.2.3 Text beneath a video image of the corresponding participant 16

7.2.3 RTT text presentation by screen readers and braille displays 16

7.2.3.1 General principles of presentation by accessibility accessories 16

7.2.3.2 Control and indications on a braille display 16

7.2.3.3 Compact presentation on a braille display 16

7.2.3.4 Screen reader with spoken output 17

8 Control of RTT in multiparty calls 17

8.1 General aspects of control of RTT 17

8.2 Control and indications of the visible user interface components 17

8.2.1 Control of the screen space 17

8.2.2 Scrolling in RTT presentation 17

8.2.3 Searching and selecting the RTT contents 17

8.2.4 Varying the RTT presentation 17

8.2.5 Notifications and status information about participants and media 17

8.2.6 Selection between meeting views 18

8.2.7 Notifications by audio and visual means 18

8.3 Control via screen readers and Braille displays 18

9 Control of calls with RTT 18

10 Transport of RTT with multiparty contents in various technical platforms 18

10.1 General transport considerations 18

10.2 Multiparty RTT transport in centralized SIP conferences 19

10.3 Multiparty RTT transport in WebRTC 19

10.4 Multiparty RTT transport in PSTN 19

11 Interaction with services 19

11.1 Services of specific importance for RTT users 19

11.2 Interpersonal conversational services 19

11.3 Conference services 20

11.4 Transforming services and relay services 20

11.5 Emergency services 20

12 Relations to specifications from other groups 21

12.1 General about support from other standardisation bodies 21

12.2 Relations to 3GPP specifications 21

12.3 Relations to GSMA specifications 21

12.3.1 Introduction 21

12.3.2 Analysis of GSMA NG.114, IR.92 and IR.94 22

12.3.3 Other GSMA profiles for conversational voice, video and RTT 23

12.4 Relations to emergency service specifications 23

13 Proposed information in EN 301 549 23

13.1 The need to update EN 301 549 to better address RTT 23

13.2 Changes to informative references in clause 2.2 24

13.3 Changes to definitions of terms in clause 3.1 24

13.4 Changes to existing requirements in clause 6. 24

13.4.1 RTT communication 24

13.4.2 Notes about RTT communication 24

13.4.3 Concurrent voice and text 25

13.4.4 Distinguishable display 25

13.4.5 Programmatically determinable send and receive origination 25

13.4.6 User identification 25

13.4.7 Visual indication of Audio with RTT 26

13.4.8 Additional clauses about RTT presentation 26

13.4.9 Interoperability 26

13.4.10 RTT responsiveness 27

13.4.11 Further requirements on RTT 27

13.5 Identification of actively speaking users 27

13.6 Identifying who wishes to communicate next 28

13.7 Awareness versus distraction 28

13.8 Programmatic determinability of the new notifications 28

13.9 Changes to requirements on services in clause 13 28

13.9.1 Access to relay services 28

13.9.2 Access to emergency services 29

Annex A: EN 301 549 extracts with proposed changes 29

A.1 Introduction 29

A.2 The EN 301 549 revisions and additions 29

2 References 29

2.2 Informative references 29

3 Definition of terms, symbols and abbreviations 30

3.1 Terms 30

6 ICT with conversational real-time voice communication 30

6.1 Audio bandwidth for speech 30

6.2 Real-Time Text (RTT) functionality 30

6.2.1 RTT provision 30

6.2.1.1 RTT communication 30

6.2.1.2 Concurrent voice and text 31

6.2.2 Display of RTT 31

6.2.2.1 Distinguishable display 31

6.2.2.2 Programmatically determinable send and receive origination 31

6.2.2.3 Speaker identification 31

6.2.2.4 Visual indicator of Audio with RTT 32

6.2.2.5 Use of RTT with Assistive Technology 32

6.2.2.6 Reviewing of received RTT 32

6.2.2.7 Presentation of relative time order of text 32

6.2.2.8 Presentation of related text from different sources 32

6.2.3 Interoperability 32

6.2.4 RTT responsiveness 33

6.2.5 Revising RTT transmitted text and text for transmission 33

6.2.6 Speed of RTT 33

6.2.7 Character representation 33

6.2.8 Media support 34

6.3 Caller ID 34

6.4 Alternatives to voice-based services 34

6.7 Activity indication 34

13 ICT providing relay or emergency service access 34

13.2 Access to relay services 34

13.3 Access to emergency services 35

13.3.1 General 35

13.3.2 Multiparty call support during emergency service calls 35

13.3.3 Callback during emergency calls 35

Annex B: Bibliography 35

Annex C: Change History 37

History 38

# Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is 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 IPR Policy, no investigation, 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.

# 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](https://portal.etsi.org/Services/editHelp%21/Howtostart/ETSIDraftingRules.aspx) (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, 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 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 Deaf and Hard of Hearing users. The inclusion of multiparty requirements will also ensure that EN 301 549 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 that would be required 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 TS 101 470 "Emergency Communications (EMTEL); Total Conversation access to Emergency Services".

[i.2] ETSI TR 103 201 "Emergency Communications (EMTEL); Total Conversation for Emergency Communication, Implementation guidelines".

[i.3] ETSI TS 103 478 "Emergency Communications (EMTEL); Pan-European Mobile Emergency Application".

[i.4] ETSI TS 103 479, "Emergency Communications (EMTEL); Core elements for network independent access to emergency services".

[i.5] ETSI TS 123 167:"Universal Mobile Telecommunications System (UMTS); LTE; IP Multimedia Subsystem (IMS) emergency sessions (3GPP TS 23.167)".

[i.6] ETSI TS 123 226 "Global Text Telephony Service Stage 2, (3GPP TS 23.226)".

[i.7] ETSI TS 124 147, "Universal Mobile Telecommunications System (UMTS); LTE; Conferencing using the IP Multimedia (IM) Core Network (CN) subsystem; Stage 3, (3GPP TS 24.147)".

[i.8] ETSI TS 124 229: "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.9] ETSI TS 124 371 "Web Real-Time Communications (WebRTC) access to the IP Multimedia (IM) Core Network (CN) subsystem (IMS); Stage 3; (3GPP TS 24.371)".

[i.10] ETSI TS 126 114: "Universal Mobile Telecommunications System (UMTS); LTE; IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction (3GPP TS 26.114)".

[i.11] ETSI TS 126 236, "Universal Mobile Telecommunications System (UMTS); LTE; Packet switched conversational multimedia applications; Transport protocols, (3GPP TS 26.236)".

[i.12] ETSI EN 301 549 (V3.2.1): "Accessibility requirements for ICT products and services".

[i.13] GSMA PRD IR.92 "IMS Profile for Voice and SMS".

[i.14] GSMA PRD IR.94 "IMS Profile for Conversational Video Service".

[i.15] GSMA PRD 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] GSMA NG.114 "IMS Profile for Voice, Video and Messaging over 5GS".

[i.18] GSMA NG.115 "IMS Profile for Voice, Video and Messaging over Untrusted WLAN Connected to 5GC".

[i.19] M. Davis, A. Phillips. "Tags for identifying languages". BCP 47, IETF September 2009.

[i.20] J. Rosenberg et.al, " SIP: Session Initiation Protocol", RFC 3261, IETF 2002.

[i.21] H. Schulzrinne et.al.," RTP: A Transport Protocol for Real-Time Applications", RFC 3550, IETF 2003.

[i.22] G. Hellstrom, P. Jones., "RTP Payload for Text Conversation". RFC 4103, IETF 2005.

[i.23] M. Davis, A. Phillips. "BCP 47 Extension T - Transformed Content". RFC 6497, IETF February 2012.

[i.24] Gellens, R., "Negotiating Human Language in Real-Time Communications", RFC 8373, IETF 2018.

[i.25] IETF RFC8825 (2021), "Overview: Real-Time Protocols for Browser-Based Applications".

[i.26] Holmberg, C. and G. Hellström, "T.140 Real-Time Text Conversation over WebRTC Data Channels", RFC 8865, IETF 2021.

[i.27] A. Began et al., "SDP Session Description Protocol", RFC 8866, IETF 2021.

[i.28] Hellström, G., "RTP-Mixer Formatting of Multiparty Real-Time Text", RFC 9071 IETF DOI 10.17487/RFC9071, (July 2021).

[i.29] ISO/IEC 10646:2020 "Information technology – Universal coded character set (UCS]".

[i.30] ITU-T Recommendation F.700."Framework Recommendation for multimedia services". ITU-T 2000.

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

[i.31] Recommendation ITU-T T.140 (1988): "Protocol for multimedia application text conversation".

[i.32] ITU-T Recommendation V.18 "Operational and interworking requirements for DCEs operating in the text telephone mode", ITU-T 1994-2000.

# 3 Definition of terms, symbols and abbreviations

## 3.1 Terms

For the purposes of the present document, the [following] terms [given in ... and the following] apply:

**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 uses "determined" where this definition uses "able to be read" (to avoid ambiguity with the word "determined").

**real-time text:** 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

**Total Conversation service:** An audiovisual conversation service providing bidirectional symmetric real-time transfer of motion video, text and voice between users in two or more locations (from [i.1]).

## 3.2 Symbols

For the purposes of the present document, the [following] symbols [given in ... and the following] apply:

## 3.3 Abbreviations

For the purposes of the present document, the [following] abbreviations [given in ... and the following] apply:

# 4 Real-time text (RTT)

## 4.1 What is RTT?

RTT, or real-time text, sends text characters input by a user (almost) immediately they are entered. This contrasts with most text chat services where the user must perform 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.

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

## 4.2 RTT is for conversation not messaging

Hearing, hard-of-hearing, and Deaf people may choose to use text messaging services as a separate service or as a supplementary service to an ongoing conference call. 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 hard-of-hearing or Deaf users 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, people 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 multi-party call becoming ineffective for most of the participants. Multiple people simultaneously using RTT could also have a similar effect.

An important factor that leads to simultaneous contributions is a lack of awareness of when other people 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 people are contributing, but they may not be able to identify who is speaking.

When several RTT users are in a multi-party 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 the RTT channel, 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 keep track of the sender’s thoughts immediately as their words are written. 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.

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 transmitted to each participant is mixed from all other participants. 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.

Speech is different from all other media described here in that the media itself does not occupy any screen area.

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 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.

### 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. Some services offer only plain text messages, while others can include images, files and video clips in the messaging channel.

### 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 it may be desirable to temporarily hide video, RTT, and messaging while watching meeting presentations. In this small screen context or other contexts, it may also be desirable to see the RTT together with the presentation but get indications when there is something new in the messaging view.

## 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 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 read 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 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.

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 audio, 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.

The text input means, e.g. keyboards, are usually the same for RTT as for text messages. Users will therefore be provided with easily operated means to control which feature will receive text input.

Controls will also be necessary to allow a participant to select which function 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 people to check back and review things contributed earlier. 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 people and with persons with speech related disabilities, even if the rapidity and lack of delays in RTT can 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. The country where she is supports calls with RTT and audio to the emergency service, so she calls the emergency number with RTT activated. 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 encapsulated in bubbles 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. Who is typing is visible 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 entries are closed and move up and eventually out of the currently presented part of the RTT area.

### 6.4.5 Deaf user participating in conference getting transcription support

A few Deaf persons participate in a large remote conference with predominantly hearing people. There are audio, video, messaging, RTT, and document presentation channels 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 in the RTT channel 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 the RTT channel. The presentation of messaging and RTT is combined in a joint real-time/messaging area where all text except the transcription is presented in real-time to make it simple for all to perceive the contributions from the typing Deaf users. The transcription is provided by automatic speech-to-text technology monitored by human text interpreters ready to jump in and transcribe when the quality of the automatic transcription is not sufficient.

An alternative way for the deaf users to contribute is to have the text interpreters read out in the audio channel what the deaf person types.

Another way is for the deaf participant to type in the RTT channel. All participants are notified that there is new content in the RTT channel, and other participants are able to read the RTT channel without viewing the transcription. The conference organisers will choose which method of fully involving their Deaf participants to use as their preferred option based on their experience of which one has worked most effectively for their group of participants in the past.

### 6.4.6 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 particular 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 in the RTT channel to the Deaf-Blind participant. Others are also allowed to contribute directly in the RTT channel.

Some meeting services may also provide human text 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 the RTT channel or the text messaging channel depending on the urgency.

### 6.4.7 Person in a critical situation making an emergency call by RTT

A hearing man gets caught up in a terrorist action. He is hidden close to the terrorist. He wants to make an emergency call but must not speak 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.

# 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 voice-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 "entry" is completed, 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 entry. 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.

### 7.1.4 Input language and character set

The users need to have the means to send text in the language and character set suitable for the call 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 milliseconds (see ITU-T F.700 [i.30]) 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 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 1.5 seconds network transmission time can then be occasionally allowed before the transmission is experienced to be severely delayed.

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 Rather loss than duplication in transport of text

The transport should guarantee that no text is duplicated. In reality, loss cannot be entirely avoided, so if the transport functions suspect that text may have been lost, that should be indicated 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

These examples are presented as potential user interface solutions. No specific, prescriptive design is required by this Technical Report or the related documents.

#### 7.2.2.1 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, separated by an end of phrase or sentence, new line, or long inactivity. Text chunks separated in that way are closed as entries and flow chronologically with other completed entries. Newly created or received text should be rendered chronologically more recently and predominantly than completed entries. If multiple participants are creating text simultaneously, received text is added in multiple insertion points in the presentation area.

#### 7.2.2.2 One column per contributing user

Another potential way of presenting multiparty RTT text 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.

#### 7.2.2.3 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 readers and braille displays

These examples are presented as potential user interface solutions. No specific, prescriptive design is required by this Technical Report or the related documents.

#### 7.2.3.1 General principles of presentation by accessibility accessories

When the text is presented by spoken output or on a braille display, it is essential that the assistive technology provide detailed control over the presentation to the user. These presentation mechanisms provide very small presentation windows and therefore need more detailed control than the visual presentation means.

#### 7.2.3.2 Control and indications on a braille display

Some Braille implementations may choose to set aside positions on the display for indications of different kinds: Indications if new text has arrived, indications of other important events in the call etc. There could also be a method of navigating to the local user’s created text and to latest read position and to latest received text. The display should in most cases not change the text contents when new text arrives, but only indicate that by a specific signal.

#### 7.2.3.3 Compact presentation on a braille display

A screen reader user with a braille display may wish to review their own text and received text simultaneously. An implementation could offer a configuration that displays the sender’s and receiver’s text in different areas of the same single row of braille cells. For example, one party’s text on the left half, and another party’s text on the right half of the braille display with an understandable division between them.

#### 7.2.3.4 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 queues to inform the user that new text has arrived.

# 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 entries 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 always 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 search the RTT text for 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 entry 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.
* 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 timeline area

### 8.2.7 Notifications by audio and visual means

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

## 8.3 Control via screen readers and Braille displays

Users of 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 these 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. 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. 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 section presents some already specified multiparty RTT transport mechanisms.

## 10.2 Multiparty RTT transport in centralized SIP conferences

For the IETF RFC 3261 "SIP" [i.20] call control protocol environment, the most commonly used transport standard for RTT is IETF RFC 4103 [i.22]. As with many other media transport protocols, IETF RFC 4103 is based on IETF RFC 3550 Real-Time Protocol RTP [i.21]. RTP has specified ways to support media transport in multiparty calls which are suitable for RTT transport by IETF RFC 4103, 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 for multiparty calls was standardized and is available as IETF RFC 9071 "RTP-Mixer Formatting of Multiparty Real-Time Text" [i.28].

IETF RFC 9071 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.

## 10.3 Multiparty RTT transport in WebRTC

A rapidly emerging real-time communication environment is WebRTC, 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 "T.140 Real-Time Text Conversation over WebRTC Data Channels" [i.31].

## 10.4 Multiparty RTT transport in PSTN

In PSTN, various text telephony protocols are specified for lower functionality transmission of text in real-time. Different protocols are, or were, used in different countries. They are in the process of being phased out, but if there is any interest in having multiparty text calls in real-time with them, then at least the protocol used in some Nordic countries, specified in ITU-T V.18 annex F [i.32] has sufficient speed and two-way simultaneous transmission so that some multi-party communication could be achieved. But no implementations have realistic performance of simultaneous voice and text, so the application would be very limited. The protocol for multi-party-unaware endpoints in IETF RFC 9071 [i.28] can be used for such purposes if the bridge is in a SIP environment and SIP/V.18 Annex F gateways are applied.

# 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 Interpersonal conversational services

Interpersonal conversational services allow calls between user devices to be set up and have a specified combination of 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 call "factory".

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.

## 11.3 Conference services

Conference services are similar to multiparty interpersonal 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 text is essential, and that all other participants become aware of new entries in the RTT channel. An alternative is that the RTT users have support from a relay service that rapidly transforms the RTT words to spoken language.

## 11.4 Transforming services and 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 sometimes called relay services. Especially when they are provided with support by national authorities in order to make electronic communications services accessible. To ease nomenclature, all variants are called relay services in the present document.

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 is established between the two call participants and a transforming entity with a three-party conference bridge with application-specific media mixing.
* Relay service support is needed in multiparty meetings and conference services to enable participation by people needing RTT communication in meetings where users predominantly use the audio channel. 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 the RTT channel, and RTT users are not prepared or capable of talking. The RTT media 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 conference bridge and connecting to the relay service with the other leg. The relay service transforms between the modalities so that the RTT channel does not need to be distributed in the conference.

The standard for indicating language is IETF BCP 47 "Tags for Identifying Languages" [i.19]. This standard can be used to support decisions about engaging relay services in calls and conferences. An extension of BCP47; IETF RFC 6497 "BCP 47 Extension-T" [i.23] 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.27] where the language of call or media can be declared, or by IETF RFC 8373 "Negotiating Human Language in Real-Time Communications" [i.24] 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".

## 11.5 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.

# 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 this report, 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 subclauses should only be seen as brief hints and advice and a starting point for assessments from experts in the standardisation bodies.

## 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 is ETSI TS 126 114 "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction." [i.10]. 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 ITU-T T.140 [i.31] on the presentation level and IETF RFC 4103 [i.22] on the transport level. IETF has recently (July 2021) published an update for IETF RFC 4103 about how to use IETF RFC 4103 in multiparty calling. The update is IETF RFC 9071 "RTP-Mixer Formatting of Multiparty Real-Time Text" [i.28]. See clause 10.2.

A way to update ETSI TS 126 114 [i.10] to support RTT multiparty calling would be to add IETF RFC 9071 for multiparty calling to where IETF RFC 4103 is mentioned, and show the negotiation for multiparty RTT support during call establishment specified in IETF RFC 9071 in the corresponding sections of ETSI TS 126 114 [i.10].

Similar modifications could be done in ETSI TS 123 226 [i.6] and ETSI TS 126 236 [i.11]. However, we find that these specifications do not seem to be maintained, so we leave it to 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.7], a paragraph could be added in 6.3.2 about support of the text (real-time text) media for multi-party 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 TS 126 114.

For WebRTC based Real Time Communications, RTT media establishment and transport of RTT is specified in IETF RFC 8865 [i.26], as described in clause 10.3 of the present document. IETF RFC 8865 already supports multiparty calling, and RFC 8865 is already referred from ETSI TS 124 371 [i.9]. The user device is expected to have ETSI TS 126 114 [i.10] 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.5]. 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.17];
* IR.92 "IMS Profile for Voice and SMS" [i.13];
* IR.94 "IMS Profile for Conversational Video Service" [i.14];
* IR.51 "IMS Profile for Voice, Video and SMS over untrusted Wi-Fi access" [i.15];
* NG.106 "IMS profile for Video, Voice and SMS over trusted Wi-Fi access" [i.16];
* NG.115 "IMS Profile for Voice, Video and Messaging over Untrusted WLAN Connected to 5GC" [i.18].

GSMA NG.114 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 and IR.92 contain RTT.

These are all profiles specifying how to use 3GPP specifications, mainly ETSI TS 126 114 (3GPP TS 26.114) [i.10] 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 sections 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.17] is titled "IMS Profile for Voice, Video and Messaging over 5GS".

IR.92 [i.13] is titled "IMS Profile for Voice and SMS"

IR.94 [i.14] 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 is for 5G networks and IR.92 for 4G LTE networks. NG.114 also specifies messaging and conversational video. IR.92 also specifies IP based SMS and has a closely related document, IR.94 specifying conversational video in LTE networks. NG.114 and the combination of IR.92 and IR.94 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 (3GPP TS 26.114) [i.10], 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 and IR.92, referring to ITU-T T.140 [i.31] and ETSI TS 126 114 [i.10] for RTT. TS 126 114 in turn refers to RFC 4103 [i.22] for initiation and transport of RTT.

Annex B.2 also adds requirements beyond what is specified in TS 26.114 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 and IR.92, 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 and IR.92 can be consulted for specification of how supplementary services shall be handled in general, 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 for RTT has recently (July 2021) got an update detailing multiparty use in IETF RFC 9071 "RTP-Mixer Formatting of Multiparty Real-Time Text" [i.28].

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

A minimum would be to add references to RFC 9071 at suitable points in Annex B.2 of IR.92 and NG.114.

IR.94 specifies how to add video to IR.92, thereby providing the full set of media for Total Conversation in LTE, similar to what NG.114 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.15],
* GSMA NG.106 "IMS profile for Video, Voice and SMS over trusted Wi-Fi access" [i.16], and
* GSMA NG.115 "IMS Profile for Voice, Video and Messaging over Untrusted WLAN Connected to 5GC" [i.18];

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 and NG.114 and specify how the situation differs from the ones specified in IR.92 [i.13] vs NG.114 [i.17]. 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 and NG.114 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 and NG.114. 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 SC 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.1], ETSI TR 103 201 [i.2], ETSI TS 103 478 [i.3] and ETSI TS 103 479 [i.4] or as EMTEL experts find appropriate. In most cases, when the documents refer to RFC 4103 for RTT initiation and transport, the modifications would add that RFC 4103 [i.22] has an update for multiparty calling in RFC 9071 [i.28]. It may also be appropriate to refer to EN 301 549 [i.12] for RTT requirements.

# 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 V.3.2.1 [i.12] 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: All references to clauses in clause 13 of the present document refer to clauses in EN 301 549 version 3.2.1 unless they are explicitly preceded by the words "in the present document".

## 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]"*

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

*"[**i.12] ETSI TS 124 229: "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)."*

## 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"*

1. 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 ITU-T F.700 [i.30], 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 RTT communication

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.

 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 Notes about 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. Two additional notes are proposed to reduce the risk that the requirement to support multiparty calling is missed, and to remind the reader that many of the accessibility requirements in other clauses are relevant for RTT implementations.

The following notes are proposed to be added to clause 6.2.1.1:

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

*NOTE 5: General accessibility requirements imply that generally available means for input and presentation of text for other applications than RTT can also be used for RTT."*

### 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. Minor rewording in Note 5 and addition of a note 6 is also proposed. In note 5, "sold" and "product" are proposed to be replaced by "provided" and "ICT system". The changed and added NOTES 1,5 and 6 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, as in a conference system, it is allowed (but not required or necessarily recommended) that floor control e.g. by use of a hand-raising tool is used to avoid confusion (in the same way that floor control is often required for those presenting/talking with voice).*

*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.*

*NOTE 6: Where ICT supports touchtone operation during voice calls, this support is expected to be maintained even when there is RTT included in the call."*

### 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 6.2.2.1 and its notes is:

*"Where ICT has RTT presentation capabilities, displayed received text from different sources and sent text shall 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 only that text from different parties is grouped per source to enable reading."*

### 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, speaker identification becomes more important also of voice users 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 speaker identification for RTT."*

### 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, all informational, navigational, and operational capabilities of the RTT interface must be available to the assistive technology.*

6.2.2.6 Reviewing of received RTT

*Where ICT has RTT presentation capabilities, the ICT shall provide the ability to review previously received RTT input.*

*NOTE: 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.*

6.2.2.7 Presentation of relative time order of text

*Where ICT has RTT presentation capabilities, the 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.*

6.2.2.8 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, 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 b) in the list in clause 6.2.3 is proposed to be modified, and item c) inserted so that the current items c) and d) move to become d) and e). Items b) and c) are proposed to read:

*"b) 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 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 124 229 [i.12] describe how IETF RFC 4103 [i.13] would apply updated by RFC 9071 [i.rfc9071],*

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

In the first line of item d) after the change, "a or b" would need to be modified to "a, b or c". In the example following the notes, "option c" should be changed to "option d".

### 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. An added Note 4 is proposed to be added to clause 6.2.4:

*"NOTE 4: The importance of this requirement is apparent in emergency service applications, where the user may lose the ability to take any further action in the user interface leaving the service agents to act on the so far entered information by the user during an emergency."*

### 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 6.2.8:

"6.2.5 Revising RTT transmitted text and text for transmission

*Where ICT utilises RTT input, the RTT user SHALL be provided with functionality to revise text by erasure from the end and to provide new input, uninfluenced by any New Line in the text.*

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 any subset thereof (including any supported emojis) supported by the ICT device or applications in use. Reception of unsupported characters shall be indicated in the way ISO/IEC 10646 provides.*

6.2.8 Media support

*When ICT has RTT support, it shall be possible to include RTT from the beginning of the call and to add or delete RTT during the call in any call leg."*

## 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 people 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 media 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."*

## 13.6 Identifying who wishes to communicate next

A "raise hand" feature is very commonly available in video 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.

## 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.12] 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.12], 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 subclause 13.3.1 "General". 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 above, and "with each media" inserted before "individually" for clarification.

The following two new clauses are also proposed to be inserted after clause 13.9.1 to clarify the most basic handling of accessible emergency calls:

"13.3.2 Multiparty call support during emergency service calls

*Multiparty calling including RTT and/or video shall be supported during emergency calls, when the initiative to add parties to the call comes from the emergency service.*

*NOTE: 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.*

13.3.3 Callback during emergency calls

*Callback from the emergency services to ICT operated by a user in an emergency service shall be supported with use of any conversational media or combinations thereof (voice, video, RTT)."*

Annex A:
EN 301 549 extracts with proposed changes

## A.1 Introduction

The present annex contains extracts from EN 301 549 v 3.2.1 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.

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 the published version of EN 301 549 [i.12].

## 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.12] ETSI TS 124 229: "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)".

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

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

NOTE 2: Where codec negotiation is implemented, other standardized codecs such as Recommendation ITU‑T G.722.2 [i.22] are sometimes used 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 allows continuous real-time conversation, the ICT shall provide functionality that allows RTT communication.

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.

NOTE 5: General accessibility requirements imply that generally available means for input and presentation of text for other applications than RTT can also be used for RTT.

#### 6.2.1.2 Concurrent voice and text

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

NOTE 1: With multiparty communication, as in a conference system, it is allowed (but not required or necessarily recommended) that floor control e.g. by use of a hand-raising tool is used to avoid confusion (in the same way that floor control is often required for those presenting/talking with voice).

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. people can be messaging in the chat field while the person is presenting/talking with RTT - in the same manner that people message using the chat feature while people 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 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 provided as a unit for the voice communication function, the local and server-side components are considered a single ICT system.

NOTE 6: Where ICT supports touchtone operation during voice calls, this support is expected to be maintained even when there is RTT included in the call.

### 6.2.2 Display of RTT

#### 6.2.2.1 Distinguishable display

Where ICT has RTT send and receive capabilities, displayed received text from different sources and sent text shall 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 only that text from different parties is grouped per source to enable reading.

#### 6.2.2.2 Programmatically determinable send and receive origination

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.

#### 6.2.2.3 Speaker identification

Where ICT has RTT capabilities, the ICT shall provide speaker identification for RTT.

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 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 that includes RTT presentation capabilities, all informational, navigational, and operational capabilities of the RTT interface must be available to the assistive technology.

#### 6.2.2.6 Reviewing of received RTT

Where ICT has RTT presentation capabilities, the ICT shall provide the ability to rewind and review previously received RTT input.

NOTE: 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.

#### 6.2.2.7 Presentation of relative time order of text

Where ICT has RTT presentation capabilities, the 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.

#### 6.2.2.8 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

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

1. 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;
2. 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 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 124 229 [i.12] describe how IETF RFC 4103 [i.13] would apply updated by RFC 9071 [i.rfc9071];
3. ICT interoperating with other ICT using WebRTC [i.rfc8825] technology to implement VoIP using RFC 8865 [i.rfc8865] to implement RTT functionality in web pages.
4. ICT interoperating with other ICT using technologies other than a, 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.
5. 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 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 options a or b to ensure that RTT is supported over the telephony connection.

### 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. 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: The importance of this requirement is apparent in emergency service applications, where the user may lose the ability to take any further action in the user interface leaving the service agents to act on the so far entered information by the user during an in emergency.

### 6.2.5 Revising RTT transmitted text and text for transmission

Where ICT utilises RTT input, the RTT user SHALL be provided with functionality to revise text by erasure from the end and to provide new input uninfluenced by any New Line in the text.

### 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 any subset thereof (including any supported emoji) supported by the ICT device or applications in use. Reception of unsupported characters shall be indicated in the way ISO/IEC 10646 provides.

### 6.2.8 Media support

When ICT has RTT support, it shall be possible to include RTT from the beginning of the call and to add or delete RTT during the call in any call leg.

## 6.3 Caller ID

Where ICT provides caller identification or similar telecommunications functions, the caller identification and similar telecommunications functions shall be available in text form as well as being programmatically determinable, unless the functionality is closed.

## 6.4 Alternatives to voice-based services

Where ICT provides real-time voice-based communication and also provides voice mail, auto-attendant, or interactive voice response facilities, the ICT shall offer users a means to access the information and carry out the tasks provided by the ICT without the use of hearing or speech.

NOTE 1: Tasks that involve both operating the interface and perceiving the information would require that both the interface and information be accessible without use of speech or hearing.

NOTE 2: Solutions capable of handling audio, RTT and video media could satisfy the above requirement.

**----------NO MODIFICATIONS PROPOSED IN CLAUSES 6.5 - 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 media 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.

**----------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 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 General

Where ICT systems support conversational real-time communication, and the system is specified for use with emergency services, access to those emergency services shall be supported 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.

### 13.3.2 Multiparty call support during emergency service calls

Multiparty calling including RTT and/or video shall be supported during emergency calls, when the initiative to add parties to the call comes from the emergency service.

NOTE: 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.

### 13.3.3 Callback during emergency calls

Callback from the emergency services to ICT operated by a user in an emergency service shall be supported with use of any conversational media or combinations thereof. (voice, video, RTT).

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

Annex B:
Bibliography

* Text telephony for all (DUST). Walter Mellors, Scott Cadzow, Edward Fitzgerald, Gunnar Hellström, HFT’06, Human Factors in Telecommunications 2006.
* 3GPP TS 22.226 Global Text Telephony Service Stage 1, 3GPP 2001 <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=627>
* 3GPP TS 22.101 Service aspects; Service principles.
<https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=605>
* ATIS-0700029 Real Time Text Mobile Device Behavior, ATIS 2017.
<https://webstore.ansi.org/standards/atis/atis0700029>
* ATIS-0700030 Real Time Text End-to-End Service Description Specification, ATIS 2018. <https://webstore.ansi.org/standards/atis/atis0700030>
* ETSI ETR 333 Text Telephony, User Requirements And Recommendations, ETSI 1997.
<https://www.etsi.org/deliver/etsi_etr/300_399/333/01_60/etr_333e01p.pdf>
* ETSI ES 202 975 V 2.1.1 Requirements for relay services, ETSI 2015
<https://www.etsi.org/deliver/etsi_es/202900_202999/202975/02.01.01_60/es_202975v020101p.pdf>
* ETSI TR 102 974 Telecommunications relay services; Background information to ES 202 975, ETSI 2009, <https://www.etsi.org/deliver/etsi_tr/102900_102999/102974/01.01.01_60/tr_102974v010101p.pdf>
* ETSI EG 202 320 Duplex Universal Speech and Text communication, ETSI 2005
<https://www.etsi.org/deliver/etsi_eg/202300_202399/202320/01.02.01_60/eg_202320v010201p.pdf>
* A. van Wijk, Ed., G. Gybels, Ed. "Framework of requirements for real-time text conversation using SIP", RFC 5194, IETF 2008.
<https://datatracker.ietf.org/doc/html/rfc5194>
* ITU-T Recommendation F.703."Multimedia conversational services". ITU-T 2000.
<https://www.itu.int/rec/T-REC-F.703-200011-I/en>
* Presentation of Text Conversation in real-time and en-bloc form, draft-hellstrom-textpreview-08, IETF 2011
<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.

Annex C:
Change History

| Date | Version | Information about changes |
| --- | --- | --- |
| 28/09/2021 | 0.0.5 | Initial draft for HF#86 meeting |
| 18/01/2022 | 0.0.6 | Draft for HF#87 |
| 18/01/2022 | 0.0.7 | Draft for HF#87 |
| 21/01/2022 | 0.0.8 | Draft for HF#87 |

# History

|  |
| --- |
| **Document history** |
| 0.0.5 | 28/09/2021 | Initial draft for HF#86 meeting |
| 0.0.6 | 18/01/2022 | Draft for HF#87 |
| 0.0.7 | 18/01/2022 | Draft for HF#87 |
| 0.0.8 | 21/01/2022 | Draft for HF#87 |
|  |  |  |

*Latest changes made on 2019-01-29*