

I n t e r n a t i o n a l T e l e c o m m u n i c a t i o n U n i o n

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1305

(07/2016)

SERIES P: TERMINALS AND SUBJECTIVE AND
OBJECTIVE ASSESSMENT METHODS

Telemeeting assessment

Effect of delays on telemeeting quality

Recommendation ITU-T P.1305

ITU-T

ITU-T P-SERIES RECOMMENDATIONS

TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	Series	P.30
		P.300
Reference systems	Series	P.40
Objective measuring apparatus	Series	P.50
		P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of speech quality	Series	P.80
Methods for objective and subjective assessment of speech and video quality	Series	P.800
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	Series	P.1100
Models and tools for quality assessment of streamed media	Series	P.1200
Telemeeting assessment	Series	P.1300
Statistical analysis, evaluation and reporting guidelines of quality measurements	Series	P.1400
Methods for objective and subjective assessment of quality of services other than speech and video	Series	P.1500

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T P.1305

Effect of delays on telemeeting quality

Summary

Recommendation ITU-T P.1305 has the primary purpose of describing the impact of echo-free transmission delays on telemeeting quality of experience (QoE). Delay has a major impact on the interaction performance of conference participants, but is not always explicitly noticed by them; its impact being commonly attributed to other participants behaviour rather than being a system feature.

Before discussing the impact of the delays, it is useful to first consider how people behave when conversing in an ideal environment. Since this has been well studied in the linguistic discipline of conversation analysis (CA), a short discussion of conversations in that context and a discussion on the impact of delay from a CA perspective is included. Some analysis based on the alternative technique of conversation surface structures is also included.

A secondary purpose of this Recommendation is to provide guidance on appropriate testing methods for evaluating the effect of delay in telemeetings.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1305	2016-07-29	12	11.1002/1000/12974

Keywords

Conversation, conversation analysis, delay, subjective test, telemeeting.

* To access the Recommendation, type the URL <http://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID. For example, <http://handle.itu.int/11.1002/1000/11830-en>.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2016

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

		Page
1	Scope.....	1
2	References.....	1
3	Definitions	2
	3.1 Terms defined elsewhere	2
	3.2 Terms defined in this Recommendation.....	3
4	Abbreviations and acronyms	3
5	Conventions	3
6	Existing Recommendations concerning the subjective quality evaluation in case of pure transmission delay	3
	6.1 Audio delay	3
	6.2 Video delay.....	3
	6.3 Synchronization between audio and video	4
7	Conversation features	4
	7.1 Two-party and multi-party conversations	4
	7.2 Two analysis techniques.....	4
	7.3 Conversation analysis and turn-taking	5
	7.4 Conversation structure and styles.....	7
	7.5 Backchannels and other behaviours during conversations	7
	7.6 Conversation surface structure analysis	7
8	The impact of delay	9
	8.1 Turn-taking with delay	9
	8.2 Conversation surface structure and delays	10
	8.3 Multiparty and two-party conversations with delay	11
	8.4 Other side-impacts of delay	11
9	Factors influencing the impact of pure delay	11
	9.1 Audiovisual telemeetings	11
	9.2 User expectations.....	12
	9.3 Asymmetric delay between participants and co-location.....	12
	9.4 Interpersonal influences.....	12
10	Testing methods.....	12
	10.1 General comments	12
	10.2 Conversational-opinion tests	13
	10.3 Group interaction tests.....	14
	10.4 Communication system effectiveness tests	16
	10.5 Micro-feature tests.....	16
11	Areas for further research regarding the impact of delay	17
	Annex A – Suggestions for free-conversation tasks	18

	Page
A.1 'Holiday' task	18
A.2 Role-playing games	18
A.3 "Who am I?" Celebrity Name-Guessing Task.....	18
A.4 Navigation tasks	19
A.5 Story with missing parts	19
A.6 Building blocks task – variation	19
A.7 Survival task – variations	20
Annex B – Suggestions for delay-critical tasks	21
B.1 Information exchange tasks	21
B.2 Random number verification task timed	21
Bibliography.....	22

Recommendation ITU-T P.1305

Effect of delays on telemeeting quality

1 Scope

This Recommendation has the primary purpose of describing the impact of echo-free transmission delays on telemeeting quality of experience (QoE). Delay is special amongst communication impairments in that it can have a major impact on how well the users interact, but very often without them being explicitly aware of its presence. Typically users comment on its impact (e.g., "we kept on interrupting each other") or assume a delayed response is due to the other participants taking a long time to respond. Consequently its impact is not always captured in conversational-opinion tests. Crucially the impact has been shown to be highly dependent on the task or activity being undertaken.

The focus in this Recommendation is therefore on the human factor – its purpose is not to define specific quantitative limits or values, discuss the sources of delay or to discuss methods of delay measurement or methods to reduce its impact.

Before discussing the impact of the delays, it is useful to first consider how people behave when conversing in an ideal environment. Since this has been well studied in the linguistic discipline of conversation analysis (CA), a short discussion of conversations in that context and a discussion on the impact of delay from a CA perspective is included. Some analysis based on the alternative technique of conversation surface structures is also included.

A secondary purpose of this Recommendation is to provide guidance on appropriate testing methods. A number of these have already been described in Annex D of [ITU-T P.1301] and will not be repeated here. However, there have been a number of new developments since that recommendation came into force and it is appropriate to expand on that knowledge.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

- [ITU-T G.114] Recommendation ITU-T G.114 (2003), *One-way transmission time*.
- [ITU-T G.131] Recommendation ITU-T G.131 (2003), *Talker echo and its control*.
- [ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.
- [ITU-T P.805] Recommendation ITU-T P.805 (2007), *Subjective evaluation of conversational quality*.
- [ITU-T P.920] Recommendation ITU-T P.920 (2000), *Interactive test methods for audiovisual communications*.
- [ITU-T P.1301] Recommendation ITU-T P.1301 (2012), *Subjective quality evaluation of audio and audiovisual multiparty telemeetings*.

- [ITU-T P.1312] Recommendation ITU-T P.1312 (2016), *Method for the measurement of the communication effectiveness of multiparty telemeetings using task performance*.
- [ITU-R BT.1359-1] Recommendation ITU-R BT.1359-1 (1998), *Relative timing of sound and vision for broadcasting*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 telemeeting [ITU-T P.1301]: A meeting in which participants are located at least two locations and the communication takes place via a telecommunication system. The term telemeeting is used to emphasize that a meeting is often more flexible and interactive than a conventional business teleconference and could also be a private meeting. The telemeeting could be audio-only, audiovisual, text-based, or a mix of these modes.

3.1.2 conversational quality [ITU-T P.1301]: The perceived quality when two or more test participants have a conversation.

3.1.3 two-party [ITU-T P.1301]: Two persons. Example: Two persons are participating in a telemeeting, having a conversation, performing a test task together, etc. If not explicitly stated differently, two-party implicates that the persons are at two locations.

3.1.4 multiparty [ITU-T P.1301]: More than two persons. Example: More than two persons are participating in a telemeeting, having a conversation, performing a test task together, etc. The term multiparty does not specify if the persons are distributed across two or more locations. If not explicitly stated differently, multiparty implicates that the persons are at two or more than two locations. When further specification is necessary, additional terms will be used (see point-to-point and multi-point) or the number of locations will be explicitly stated.

3.1.5 point-to-point [ITU-T P.1301]: Two locations. Example: A multiparty point-to-point telemeeting means that more than two interlocutors are taking part, and the interlocutors are at exactly two locations. That means that in one location more than one interlocutors are present, as there are more than two persons.

3.1.6 multi-point [ITU-T P.1301]: More than two locations. Example: A multiparty multi-point telemeeting means that more than two interlocutors are taking part, and the interlocutors are located across more than two locations. Multi-point does not specify if one or more than one interlocutor may be present at each location. In the special case that only one person is present at each location, the term one-per-site will be used.

3.1.7 one-per-site [ITU-T P.1301]: One person per connected location. Example: In a multiparty one-per-site telemeeting more than two sites are connected with only one person present at each site.

3.1.8 conversation analysis (CA) [b-Sacks, 1974]: The study of social interaction, in particular focusing on conversations.

3.1.9 turn construction unit (TCU) [b-Sacks, 1974]: A conversation analysis term describing the fundamental segment of speech in a conversation – essentially a piece of speech that constitutes an entire 'turn'.

3.1.10 transition relevance place (TRP) [b-Sacks, 1974]: A conversation analysis term which indicates where a turn or floor exchange can take place between speakers.

3.1.11 system effectiveness [ITU-T P.1312]: The ratio, expressed as a percentage, between the performance score achieved on a given system and the performance score obtained in face-to-face communication.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 choral behaviour: Many people doing the same thing at the same time.

3.2.2 phatic expression: An expression whose function is part of social interaction rather than to convey information.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

CA	Conversation Analysis
MOS	Mean Opinion Score
QoE	Quality of Experience
QoS	Quality of Service
TCU	Turn Constructional Unit
TRP	Transition Relevance Place

5 Conventions

None.

6 Existing Recommendations concerning the subjective quality evaluation in case of pure transmission delay

In the following existing Recommendations on audio delay, video delay and audio-video synchronization the current understanding is outlined. These Recommendations have been developed for two-party interaction only. The multi-party case needs further study to ensure their applicability.

6.1 Audio delay

Recommendation [ITU-T G.114] states that,

"Regardless of the type of application, it is recommended to not exceed a one-way delay of 400 ms for general network planning..."

"Although a few applications may be slightly affected by end-to-end (i.e., "mouth-to-ear" in the case of speech) delays of less than 150 ms, if delays can be kept below this figure, most applications, both speech and non-speech, will experience essentially transparent interactivity."

The delay values in [ITU-T G.114] are based on the effect of pure audio delay only, i.e., in the complete absence of any echo. It is well-known that if echo is present, longer audio delays become easier detectable and much more disturbing. Effects of talker echo at different one-way echo-path delays are further described in [ITU-T G.131].

6.2 Video delay

Most investigations regarding quality impact of delay have been performed regarding audio delays, but in an earlier version of [ITU-T G.114] (02/96), in clause B.2.3, it is mentioned that:

"Tests were performed to assess the interaction between delay and user applications. In these tests a comparison of telephone conversations with videophone were made and it was shown that there is little difference between both types of connection."

This result was obtained for synchronized audio and video. Some research has been performed looking at pure audio-visual delay in an interactive setting, but there is a need for more research to clearly confirm the recommendations of [ITU-T G.114] for this case, especially regarding multi-party communication.

6.3 Synchronization between audio and video

The limits within which audio and video are perceived to be synchronized are well examined, at least for TV screens. According to [ITU-R BT.1359-1],

"subjective evaluations show that detectability thresholds are about +45 ms to –125 ms and acceptability thresholds are about +90 ms to –185 ms on the average, a positive value indicates that sound is advanced with respect to vision".

As video encoding usually requires more time than audio coding, audio should generally be delayed to obtain audio-video synchronization [b-Berndtsson, 2012]. It is likely that it is easier to identify audio-video asynchrony as a technical issue than synchronous delay. Following from this, audio-video asynchrony will most likely be reflected more critically in mean opinion score (MOS) ratings.

Low video frame rates and video frame rate jitter can influence the impression of audio-video synchronization.

7 Conversation features

7.1 Two-party and multi-party conversations

There are differences between two-way conversations and multi-party conversations. For instance in multi-party conversations:

- participants are likely to spend a greater proportion of the time listening rather than talking. Usually this will vary as the discussion proceeds, with talkers being more active or passive at different times;
- telemeeting systems do not always attempt to reproduce the exact signals from the sending end at the receiving end; system designers are usually more interested in creating an environment where participants can interact effectively, and this can involve multiple technologies such as spatial audio, level compensation, noise reduction etc.;
- participant interaction is more complicated: in a two-party conversation the role of the speaker and the listener simply alternate, whereas in a multi-party conversation, the next speaker could be any one of several current listeners;
- passive participants are not only listening to what the active participants are saying, they are listening to the active participants interacting.

7.2 Two analysis techniques

In this Recommendation two different techniques for analysing a conversation are discussed. The first is based on the linguistic discipline of conversation analysis (CA) [b-Sacks, 1974] and focuses on analysing how the participants interact. For evaluating the impact of delay, the two most relevant analysis concepts from CA are turn-taking (see clause 7.3) and backchannels (see clause 7.6).

The second technique analyses patterns of the temporal occurrence of utterances without labelling their purpose. This approach does not rely on the content of the utterances and can therefore be