

Recommendation

ITU-T P.836 (03/2023)

SERIES P: Telephone transmission quality, telephone installations, local line networks

Methods for objective and subjective assessment of speech and video quality

Simulating conversations for the prediction of speech quality

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	P.10–P.19
Voice terminal characteristics	P.30–P.39
Reference systems	P.40–P.49
Objective measuring apparatus	P.50–P.59
Objective electro-acoustical measurements	P.60–P.69
Measurements related to speech loudness	P.70–P.79
Methods for objective and subjective assessment of speech quality	P.80–P.89
Voice terminal characteristics	P.300–P.399
Objective measuring apparatus	P.500–P.599
Measurements related to speech loudness	P.700–P.709
Methods for objective and subjective assessment of speech and video quality	P.800–P.899
Audiovisual quality in multimedia services	P.900–P.999
Transmission performance and QoS aspects of IP end-points	P.1000–P.1099
Communications involving vehicles	P.1100–P.1199
Models and tools for quality assessment of streamed media	P.1200–P.1299
Telemeeting assessment	P.1300–P.1399
Statistical analysis, evaluation and reporting guidelines of quality measurements	P.1400–P.1499
Methods for objective and subjective assessment of quality of services other than speech and video	P.1500–P.1599

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

Recommendation ITU-T P.836

Simulating conversations for the prediction of speech quality

Summary

Recommendation ITU-T P.836 provides a conversation simulation model which is able to simulate realistic conversational behaviour to produce conversations on the semantic level, as well as on the speech signal level. The simulation can replicate conversations with different interactivity patterns and the resulting simulated conversations will reflect changes in this conversation behaviour due to delayed transmission and packet loss. The simulated conversations may be used to predict conversational quality in combination with signal-based or parametric quality prediction models, such as the E-model, e.g., in drive-test scenarios.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.836	2023-03-01	12	11.1002/1000/15470

Keywords

Conversation simulation, conversational quality prediction, speech quality.

* 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/11.1002/1000/11470>.

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 received notice of intellectual property, protected by patents/software copyrights, 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 appropriate ITU-T databases available via the ITU-T website at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2023

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	1
	3.1 Terms defined elsewhere	1
	3.2 Terms defined in this Recommendation	2
4	Abbreviations and acronyms	2
5	Conventions	2
6	Conversational behaviour	2
	6.1 Dialogue acts and concepts.....	3
	6.2 Transcriptions	3
	6.3 Conversation disruptions.....	4
	6.4 Turn-taking	4
	6.5 Conversational interactivity	4
7	Simulation of conversational behaviour	5
	7.1 Incremental simulation framework	5
	7.2 Turn-taking dialogue manager.....	8
	7.3 Turn-taking	9
	7.4 Conversation disruptions.....	10
8	Application to the conversational quality prediction	10
	8.1 Quality prediction with the E-model	10
	Annex A – Python source code.....	11
	Appendix I – Example agenda of simulated agents.....	12
	Appendix II – Quality predictions from simulated conversations with the E-model	14
	Bibliography	15

Python reference implementation of the conversation simulation.

Recommendation ITU-T P.836

Simulating conversations for the prediction of speech quality

1 Scope

This Recommendation¹ describes a methodology on how to create a conversation simulation environment that is able to produce conversations that replicate certain characteristics of similar conversations recorded in a conversation test as described in [ITU-T P.800], [ITU-T P.805] and [ITU-T P.804]. The parameters aimed to replicate with the simulation approach are the turn-taking behaviours, the lengths of talk spurts, the interactivity of the conversation, as well as the intents conveyed in each utterance. In contrast to artificial conversational speech described in [ITU-T P.59], the simulation replicates the interactivity patterns of concrete conversation scenarios (e.g., conversation scenarios of [ITU-T P.805] Appendices IV-IX), as well as the impact of degradations such as transmission delay and packet loss on the conversation. The simulated conversations may be used in combination with existing quality models to predict the conversational quality but in itself does not predict the conversational quality.

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.107] Recommendation ITU-T G.107 (2015), *The E-model: a computational model for use in transmission planning*.
- [ITU-T G.107.2] Recommendation ITU-T G.107.2 (2023), *Fullband E-model*.
- [ITU-T P.59] Recommendation ITU-T P.59 (1993), *Artificial conversational speech*.
- [ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.
- [ITU-T P.804] Recommendation ITU-T P.804 (2017), *Subjective diagnostic test method for conversational speech quality analysis*.
- [ITU-T P.805] Recommendation ITU-T P.805 (2007), *Subjective evaluation of conversational quality*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 double talk (DT) [b-ITU-T P.832]: When near-end and far-end speech occur simultaneously at a given point, typically the terminal under test.

¹ This Recommendation includes an electronic attachment with a reference implementation of the conversation simulation in Python (Annex A).

3.1.2 conversation test [b-ITU-T P.832]: A subjective test in which two participants have a real-time conversation, as described in Annex A to [ITU-T P.800] and in [b-Telephony].

3.2 Terms defined in this Recommendation

This Recommendation defines the following term:

3.2.1 conversation disruption: A turn in a conversation that explicitly or implicitly asks for the retransmission of previous information.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ASR	Automatic Speech Recognition
CD	Conversation Disruption
DT	Double Talk
IU	Incremental Unit
MOS	Mean Opinion Score
MS	Mutual Silence
NLG	Natural Language Generation
NLU	Natural Language Understanding
P-CA	Parametric Conversation Analysis
POTS	Plain Old Telephone System
RNV	Random Number Verification
SA	Speaker A
SAR	Speaker Alternation Rate
SARc	Corrected Speaker Alternation Rate
SB	Speaker B
SCT	Short Conversation Test
TTS	Text-to-Speech
VoIP	Voice over Internet Protocol

5 Conventions

None.

6 Conversational behaviour

For the simulation to be used to predict the speech quality in telephone² conversations, a set of conversational behaviour is defined that can be extracted from real conversations and will be modelled in the simulation. This behaviour includes temporal parameters in the form of the timing of turns and measures of the interactivity of the conversation, as well as semantic representation in the form of dialogue acts and concepts that encapsulate the meaning behind the speaking turns. Even

² "Telephone" here refers to any form of remote speech communication like plain old telephone system (POTS), mobile communication, or over-the-top voice over internet protocol (VoIP) services.