

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

Series P

Supplement 25

(01/2011)

SERIES P: TERMINALS AND SUBJECTIVE AND
OBJECTIVE ASSESSMENT METHODS

Parameters describing the interaction with multimodal dialogue systems

ITU-T P-series Recommendations – Supplement 25



International
Telecommunication
Union

Document Sharing Hub

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
		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

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

Supplement 25 to ITU-T P-series Recommendations

Parameters describing the interaction with multimodal dialogue systems

Summary

Supplement 25 to the ITU-T P-series Recommendations provides definitions for a set of parameters which can be extracted from services which rely on multimodal dialogue systems. The parameters can be extracted from logged (test) user interactions with the service under consideration. They quantify the flow of the interaction, the behaviour of the user and the system, and the performance of the speech technology devices involved in the interaction. They provide useful information for system development, optimization and maintenance, and are complementary to subjective quality judgments. The list is an amendment and extension of the respective list of parameters for speech-based services which is given in Supplement 24 to the ITU-T P-series Recommendations.

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T P Suppl. 25	2011-01-27	12

Keywords

Assessment, automatic speech recognition, automatic speech understanding, dialogue management, gesture recognition, interaction parameter, multimodal dialogue system, speech generation.

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

Compliance with this publication is voluntary. However, the publication may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the publication 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 publication is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this publication 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 publication development process.

As of the date of approval of this publication, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this publication. 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 2011

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	4
4	Abbreviations.....	5
5	Conventions	5
6	Introduction	6
7	Characteristics of interaction parameters.....	7
8	Review of interaction parameters	7
8.1	Dialogue- and communication-related parameters.....	8
8.2	Meta-communication-related parameters	10
8.3	Cooperativity-related parameters	13
8.4	Task-related parameters.....	14
8.5	Input-related parameters	15
8.6	Output-related parameters	19
8.7	Further parameters.....	20
9	Interpretation of interaction parameter values	20

Copyright International Telecommunication Union/ITU Telecommunication Sector
 Provided by IHS Markit under license with ITU-T
 No reproduction or networking permitted without license from IHS
 Full document details: ITU-T SERIES P SUPP 25 2011

Supplement 25 to ITU-T P-series Recommendations

Parameters describing the interaction with multimodal dialogue systems

1 Scope

This supplement describes parameters providing information on the interaction with services which are based on multimodal dialogue systems, as seen by the system developer and service operator. Multimodal dialogue systems addressed by this supplement enable a multimodal interaction with a human user. Such systems offer one or more modalities for input (e.g., speech, gesture, touch) and one or more output modalities (e.g., a graphical user interface, spoken output, an embodied conversational agent) and may have automatic speech, gesture or touch recognition, speech understanding, a fusion module, dialogue management, response generation, a fission module and speech, graphical or audiovisual output capabilities. They may provide access to information stored in a database, or allow different types of transactions to be performed, and they are frequently offered on smart-phone platforms.

The parameters defined here quantify the flow of the interaction, the behaviour of the user and the system, and the performance of the devices involved in the interaction. For extracting all parameters, the multimodal dialogue system has to be accessible as a glass box; still, some parameters may also be extracted in a black-box approach, i.e., without access to the individual system components. The extraction can partially be performed automatically, and partially relies on a human expert transcribing and annotating interaction log files. The parameters address system performance from a system developer's point-of-view; thus, they provide complementary information to subjective evaluation experiments. Further guidance on subjective evaluation methods in general and on the assessment of speech output devices and spoken dialogue systems is available in [ITU-T P.800], [ITU-T P.85] and [ITU-T P.851], and in the Handbook on Telephonometry. This guidance, however, does not yet cover multimodal systems.

2 References

- [ITU-T P.10] Recommendation ITU-T P.10/G.100 (2006), *Vocabulary for performance and quality of service*.
- [ITU-T P.85] Recommendation ITU-T P.85 (1994), *A method for subjective performance assessment of the quality of speech voice output devices*.
- [ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.
- [ITU-T P.851] Recommendation ITU-T P.851 (2003), *Subjective quality evaluation of telephone services based on spoken dialogue systems*.
- [ITU-T Handbook] ITU-T Handbook on Telephonometry (1993).
- [IEC 60268-16] IEC Standard 60268-16 (1998), *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index*.
- [Bernsen-1] Bernsen, N.O., Dybkjær, H., Dybkjær, L. (1998), *Designing interactive speech systems: From first ideas to user testing*. Springer, DE-Berlin.
- [Bernsen-2] Bernsen, N.O. (2002), *Multimodality in Language and Speech Systems – From Theory to Design Support Tool*. In Granström, B., House, D., and Karlsson, I. (Eds.): *Multimodality in Language and Speech Systems*, Dordrecht, Kluwer Academic Publishers, 93-148.

- [Billi] Billi, R., Castagneri, G., Danieli, M. (1996), *Field trial evaluations of two different information inquiry systems*. In: Proc. 3rd IEEE Workshop on Interactive Voice Technology for Telecommunications Applications (IVTTA'96), US-Basking Ridge NJ, 129-134.
- [Boros] Boros, M., Eckert, W., Gallwitz, F., Gorz, G., Hanrieder, G., Niemann, H. (1996), *Towards understanding spontaneous speech: Word accuracy vs. concept accuracy*. In: Proc. 4th Int. Conf. on Spoken Language Processing (ICSLP'96), IEEE, US-Piscataway NJ, 2, 1009-1012.
- [Carletta] Carletta, J. (1996), *Assessing agreement of classification tasks: The kappa statistics*. Computational Linguistics, 22(2), 249-254.
- [Chu] Chu, M., Peng, H. (2001), *An objective measure for estimating MOS of synthesized speech*. In: Proc. 7th Europ. Conf. on Speech Communication and Technology (Eurospeech 2001 – Scandinavia), DK-Aalborg, 3, 2087-2090.
- [Cookson] Cookson, S. (1988), *Final evaluation of VODIS – Voice operated data inquiry system*. In: Proc. of Speech'88, 7th FASE Symposium, UK-Edinburgh, 4, 1311-1320.
- [Danieli] Danieli, M., Gerbino, E. (1995), *Metrics for evaluating dialogue strategies in a spoken language system*. In: Empirical Methods in Discourse Interpretation and Generation. Papers from the 1995 AAAI Symposium, US-Stanford CA, AAAI Press, US-Menlo Park CA, 34-39.
- [Fraser] Fraser, N. (1997), *Assessment of interactive systems*. In: Handbook on Standards and Resources for Spoken Language Systems (D. Gibbon, R. Moore and R. Winski, eds.), Mouton de Gruyter, DE-Berlin, 564-615.
- [Gerbino] Gerbino, E., Baggia, P., Ciaramella, A., Rullent, C. (1993), *Test and evaluation of a spoken dialogue system*. In: Proc. Int. Conf. Acoustics Speech and Signal Processing (ICASSP'93), IEEE, US-Piscataway NJ, 2, 135-138.
- [Gibbon] Gibbon, D., Moore, R., Winski, R., Eds. (2000), *Handbook on Standards and Resources for Spoken Language Systems*. Mouton de Gruyter, DE-Berlin.
- [Glass] Glass, J., Polifroni, J., Seneff, S., Zue, V. (2000), *Data collection and performance evaluation of spoken dialogue systems: The MIT experience*. In: Proc. 6th Int. Conf. on Spoken Language Processing (ICSLP 2000), CN-Beijing, 4, 1-4.
- [Goodine] Goodine, D., Hirschman, L., Polifroni, J., Seneff, S., Zue, V. (1992), *Evaluating interactive spoken language systems*. In: Proc. 2nd Int. Conf. on Spoken Language Processing (ICSLP'92), CA-Banff, 1, 201-204.
- [Grice] Grice, H.P. (1975), *Logic and conversation*. In: Syntax and Semantics, Vol. 3: Speech Acts (P. Cole and J.L. Morgan, eds.), Academic Press, US-New York NY, 41-58.
- [Hirschman] Hirschman, L., Pao, C. (1993), *The cost of errors in a spoken language system*. In: Proc. 3rd Europ. Conf. on Speech Communication and Technology (Eurospeech'93), DE-Berlin, 2, 1419-1422.
- [Kamm] Kamm, C.A., Litman, D.J., Walker, M.A. (1998), *From novice to expert: The effect of tutorials on user expertise with spoken dialogue systems*. In: Proc. 5th Int. Conf. on Spoken Language Processing (ICSLP'98), AU-Sydney, 4, 1211-1214.

- [Kühnel] Kühnel, C., Weiss, B., Möller, S., (2010), *Parameters describing multimodal interaction – Definitions and three usage scenarios*. Procedures of the 11th Annual Conference of the ISCA (Interspeech 2010), JP – Tokyo 2014-2017.
- [Möller] Möller, S. (2005), *Quality of telephone-based spoken dialogue systems*. Springer, US-New York NY.
- [Nigay] Nigay L. and Coutaz J. (1993), *A design space for multimodal systems: concurrent processing and data fusion*. In: Proc. Of the INTERACT and CHI, 172-178.
- [NIST SRST] NIST Speech Recognition Scoring Toolkit (2001), *Speech recognition scoring toolkit*. National Institute of Standards and technology, <http://www.nist.gov/speech/tools>, US-Gaithersburg MD.
- [Perakakis] Perakakis, M and Potamianos, A. (2008), *Multimodal system evaluation using modality efficiency and synergy metrics*. In: *Proc. of the 10th international conference on Multimodal interfaces (ICMI '08)*. ACM, New York, NY, USA, 9-16.
- [Picone-1] Picone, J., Doddington, G.R., Pallett, D.S. (1990), *Phone-mediated word alignment for speech recognition evaluation*. IEEE Trans. Acoustics Speech and Signal Processing, 38(3), 559-562.
- [Picone-2] Picone, J., Goudie-Marshall, K.M., Doddington, G.R., Fisher, W. (1986), *Automatic text alignment for speech system evaluation*. IEEE Trans. Acoustics Speech and Signal Processing, 34(4), 780-784.
- [Polifroni] Polifroni, J., Hirschman, L., Seneff, S., Zue, V. (1992), *Experiments in evaluating interactive spoken language systems*. In: Proc. DARPA Speech and Natural Language Workshop, US-Harriman CA, 28-33.
- [Price] Price, P.J., Hirschman, L., Shriberg, E., Wade, E. (1992), *Subject-based evaluation measures for interactive spoken language systems*. In: Proc. DARPA Speech and Natural Language Workshop, US-Harriman CA, 34-39.
- [San-Segundo] San-Segundo, R., Montero, J.M., Colás, J., Gutiérrez, J., Ramos, J.M., Pardo, J.M. (2001), *Methodology for dialogue design in telephone-based spoken dialogue systems: A Spanish train information system*. In: Proc. 7th Europ. Conf. on Speech Communication and Technology (Eurospeech 2001 – Scandinavia), DK-Aalborg, 3, 2165-2168.
- [Simpson] Simpson, A., Fraser, N.M. (1993), *Black box and glass box evaluation of the SUNDIAL system*. In: Proc. 3rd Europ. Conf. on Speech Communication and Technology (Eurospeech'93), DE-Berlin, 2, 1423-1426.
- [Skowronek] Skowronek, J. (2002), *Entwicklung von Modellierungsansätzen zur Vorhersage der Dienstqualität bei der Interaktion mit einem natürlichsprachlichen Dialogsystem*. Diploma thesis (unpublished), Institut für Kommunikationsakustik, Ruhr-Universität, DE-Bochum.
- [Strik-1] Strik, H., Cucchiari, C., Kessens, J.M. (2001), *Comparing the performance of two CSRs: How to determine the significance level of the differences*. In: Proc. 7th Europ. Conf. on Speech Communication and Technology (Eurospeech 2001 – Scandinavia), DK-Aalborg, 3, 2091-2094.
- [Strik-2] Strik, H., Cucchiari, C., Kessens, J.M. (2000), *Comparing the recognition performance of CSRs: In search of an adequate metric and statistical significance test*. In: Proc. 6th Int. Conf. on Spoken Language Processing (ICSLP 2000), CN-Beijing, 4, 740-743.