International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Series P

Supplement 24

(10/2005)

SERIES P: TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Parameters describing the interaction with spoken dialogue systems

ITU-T P-series Recommendations – Supplement 24

Copyright International Telecommunication Union/ITU Telcommunication 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 24 2005

Licensee=/, User=,
Not for Resale,

International Telecommunication Union
Documents
Document Sharing Hub

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	Series	P.10
Subscribers' lines and sets	Series	P.30
		P.300
Transmission standards	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 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

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

--`,,```,,,,````-`-`,,`,,`,`,,`---

Documents
Document Sharing Hub

Copyright International Telecommunication Union/ITU Telcommunication Sector
Provided by IHS Markit under license with ITU-T
No reproduction or networking permitted without license from IHS

Licensee=/, User=,
Not for Resale,

Full document details: ITU-T SERIES P SUPP 24 2005

Supplement 24 to ITU-T P-series Recommendations

Parameters describing the interaction with spoken dialogue systems

Summary

This Supplement provides definitions for a set of parameters which can be extracted from services which rely on spoken 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 collected according to ITU-T Rec. P.851.

Source

Supplement 24 to ITU-T P-series Recommendations was agreed on 21 October 2005 by ITU-T Study Group 12 (2005-2008).

Keywords

Assessment, automatic speech recognition, automatic speech understanding, dialogue management, interaction parameter, speech generation, speech technology, spoken dialogue system.

Documents
Document Sharing Hub

Copyright International Telecommunication Union/ITU Telcommunication Sector
Provided by IHS Markit under license with ITU-T
No reproduction or networking permitted without license from IHS
Licensee=/, User=,
Not for Resale,
Full document details: ITU-T SERIES P SUPP 24 2005

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. 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, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

© ITU 2005

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

Documents
Document Sharing Hub

Copyright International Telecommunication Union/ITU Telcommunication Sector
Provided by IHS Markit under license with ITU-T
No reproduction or networking permitted without license from IHS
Licensee=/, User=,
Not for Resale,
Full document details: ITU-T SERIES P SUPP 24 2005

CONTENTS

		Page
1	Scope	1
2	References	1
3	Definitions	1
4	Abbreviations	2
5	Introduction	2
6	Characteristics of interaction parameters	3
7	Review of interaction parameters	4
8	Interpretation of interaction parameter values	14
BIBLIOGRAPHY		16

Documents
Document Sharing Hub
Copyright International Telecommunication Union/ITU Telcommunication Sector
Provided by IHS Markit under license with ITU-T
No reproduction or networking permitted without license from IHS
Licensee=/, User=,
Not for Resale,
Full document details: ITU-T SERIES P SUPP 24 2005

--`,,```,,,,````-`-`,,`,,`,`,,`---

Documents
Document Sharing Hub

Copyright International Telecommunication Union/ITU Telcommunication Sector
Provided by IHS Markit under license with ITU-T
No reproduction or networking permitted without license from IHS

Licensee=/, User=,
Not for Resale,

Full document details: ITU-T SERIES P SUPP 24 2005

Supplement 24 to ITU-T P-series Recommendations

Parameters describing the interaction with spoken dialogue systems

1 Scope

This Supplement describes parameters providing information on the interaction with services which are based on spoken dialogue systems, as seen by the system developer and service operator. Spoken dialogue systems addressed by this Supplement enable a spoken language interaction with a human user via the telephone network on a turn-by-turn basis, and have automatic speech recognition, speech understanding, dialogue management, response generation, and speech output capabilities. They may provide access to information stored in a database, or allow different types of transactions to be performed.

The parameters defined here 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. For extracting all parameters, the spoken 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 with spoken dialogue systems for which recommendations are given in ITU-T Rec. P.851. Further guidance on subjective evaluation methods in general and on the assessment of speech output devices, is available in ITU-T Recs P.800 and P.85, and in the Handbook on Telephonometry. The parameters listed in this Supplement do not specifically refer to possible degradations introduced by the transmission channel. These effects are an item for further study by ITU-T SG 12.

2 References

– ITU-T Recommendation P.85 (1994), *A method for subjective performance assessment of the quality of speech voice output devices*.

– ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*.

– ITU-T Recommendation P.851 (2003), *Subjective quality evaluation of telephone services based on spoken dialogue systems.*

– ITU-T *Handbook on Telephonometry* (1992).

3 Definitions

For definitions not listed here, please refer to ITU-T Rec. P.10.

3.1 barge-in: The ability of a human to speak over a system prompt or system output [10].

3.2 dialogue: A conversation or an exchange of information. As an evaluation unit: One of several possible paths through the dialogue structure.

3.3 efficiency: Measures of the accuracy and completeness of system tasks relative to the resources (e.g., time, human effort) used to achieve the specific system tasks.

3.4 exchange: A pair of contiguous and related turns, one spoken by each party in the dialogue [8].

3.5 functionality: Capability of the system to provide functions which meet stated and implied needs when the system is used under specific conditions.

Documents
Document Sharing Hub
Copyright International Telecommunication Union/ITU Telcommunication Sector
Provided by IHS Markit under license with ITU-T
No reproduction or networking permitted without license from IHS
Licensee=/, User=,
Not for Resale,
Full document details: ITU-T SERIES P SUPP 24 2005

3.6 **meta-communication**: The communication about communication, e.g., for resolving misunderstandings ("Did I understand you right?") or for reaching agreement on the use of the language.

3.7 **performance**: The ability of a unit to provide the function it has been designed for.

3.8 **speech technology**: The discipline concerned with the research and development of spoken language input and output systems, using contributions from the neighbouring disciplines of acoustics, electrical engineering, statistics, phonetics, natural language processing, and involving system requirements specification, design, implementation and evaluation, corpus and linguistic resource processing, and consumer oriented product evaluation [10].

3.9 **spoken dialogue system**: A computer system with which human users interact via spoken language on a turn-by-turn basis.

3.10 **task**: All the activities which a user must develop in order to attain a fixed objective in some domain.

3.11 **task-oriented dialogue**: A dialogue concerning a specific subject, aiming at an explicit goal (such as resolving a problem or obtaining specific information) [8].

3.12 **transaction**: The part of a dialogue devoted to a single high-level task (e.g., making a travel booking or checking a bank account balance). A transaction may be coextensive with a dialogue, or a dialogue may consist of more than one transaction [8].

3.13 **turn**: Utterance. A stretch of speech, spoken by one party in a dialogue, from when this party starts speaking until another party definitely takes over [1].

3.14 **utterance**: See turn.

4 Abbreviations

ASR	Automatic Speech Recognition
AVM	Attribute-Value Matrix
AVP	Attribute-Value Pair
DARPA	Defense Advanced Research Projects Agency
DP	Dynamic Programming
DTMF	Dual Tone Multiple Frequency
IVR	Interactive Voice Response
MOS	Mean Opinion Score
SDS	Spoken Dialogue System
WoZ	Wizard-of-Oz

5 Introduction

Spoken dialogue systems (SDSs), i.e., computer systems with which human users interact via spoken language on a turn-by-turn basis, may be part of modern telephone networks. They enable access to databases and transactions via the telephone, e.g., for obtaining train or airline timetable information, stock exchange rates, tourist information, or to perform bank account operations or make hotel reservations. In contrast to simple interactive voice response (IVR) systems with DTMF input, SDSs offer the full range of speech interaction capabilities, including the recognition of user speech, the assignment of meaning to the recognized words, the decision on how to continue the

--`,,```,,,,````-`-`,,`,,`,`,,`---
Documents
Document Sharing Hub
Copyright International Telecommunication Union/ITU Telcommunication Sector
Provided by IHS Markit under license with ITU-T
No reproduction or networking permitted without license from IHS
Licensee=/, User=,
Not for Resale,
Full document details: ITU-T SERIES P SUPP 24 2005