

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

F.748.15

(03/2022)

SERIES F: NON-TELEPHONE TELECOMMUNICATION
SERVICES

Multimedia services

Framework and metrics for digital human application systems

Recommendation ITU-T F.748.15

ITU-T

ITU-T F-SERIES RECOMMENDATIONS
NON-TELEPHONE TELECOMMUNICATION SERVICES

TELEGRAPH SERVICE	
Operating methods for the international public telegram service	F.1–F.19
The gentex network	F.20–F.29
Message switching	F.30–F.39
The international telemesssage service	F.40–F.58
The international telex service	F.59–F.89
Statistics and publications on international telegraph services	F.90–F.99
Scheduled and leased communication services	F.100–F.104
Phototelegraph service	F.105–F.109
MOBILE SERVICE	
Mobile services and multideestination satellite services	F.110–F.159
TELEMATIC SERVICES	
Public facsimile service	F.160–F.199
Teletex service	F.200–F.299
Videotex service	F.300–F.349
General provisions for telematic services	F.350–F.399
MESSAGE HANDLING SERVICES	F.400–F.499
DIRECTORY SERVICES	F.500–F.549
DOCUMENT COMMUNICATION	
Document communication	F.550–F.579
Programming communication interfaces	F.580–F.599
DATA TRANSMISSION SERVICES	F.600–F.699
MULTIMEDIA SERVICES	F.700–F.799
ISDN SERVICES	F.800–F.849
UNIVERSAL PERSONAL TELECOMMUNICATION	F.850–F.899
ACCESSIBILITY AND HUMAN FACTORS	F.900–F.999

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

Recommendation ITU-T F.748.15

Framework and metrics for digital human application systems

Summary

Recommendation ITU-T F.748.15 specifies a framework for digital human application systems, and proposes corresponding subjective and objective metrics for the dimensions of image, speech, animation, interactive processing and multimodal input/output. This Recommendation can be used to guide relevant parties to test, select or evaluate a digital human application system. The metrics can reflect the current state of the digital human application system by providing meaningful comparison dimensions.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T F.748.15	2022-03-16	16	11.1002/1000/14962

Keywords

Animation module, digital human, framework and metrics, image module, interactive processing module, multimodal input module, multimodal output module, speech module.

* 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/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 2022

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	Framework for digital human application system	2
7	Metrics and normative description	3
	7.1 Overview	3
	7.2 Image	4
	7.3 Speech.....	5
	7.4 Animation	6
	7.5 Interactive processing.....	8
	7.6 Multimodal input	12
	7.7 Multimodal output	12
	Appendix I – Use case for digital human application system.....	14
	I.1 Virtual customer service.....	14
	Bibliography.....	16

Recommendation ITU-T F.748.15

Framework and metrics for digital human application systems

1 Scope

This Recommendation specifies a framework and metrics for digital human (i.e., intelligent-driven digital human) application systems, and addresses:

- a) a framework;
- b) metrics for an image module;
- c) metrics for a speech module;
- d) metrics for an animation module;
- e) metrics for an interactive processing module;
- f) metrics for a multimodal input module;
- g) metrics for a multimodal output module.

This Recommendation applies to quality evaluation for digital human application systems.

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.

None.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 application system [b-ISO/IEC TR 10032]: A collection of application processes which utilizes the services provided by the human-computer interface, communications facility, and data management system to perform the processing necessary to meet the requirements of the information system.

3.1.2 text-to-speech synthesis (TTS) [b-ITU-T P.10]: A TTS process generates a speech signal from text codes. It is usually composed of the two parts:

- a language-dependent text processing part (the high level processing part), which generates from the character string (by reading rules, vocabulary and semantic analysis) a set of phonetic, prosodic, etc., parameters which are used by:
- an acoustical signal generating part, the synthesizer itself, which generates the audible speech.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 digital human: A computer application that integrates the technologies of computer graphics, computer vision, intelligent speech and natural language processing. It can be used for digital content generation and human-computer interaction to help improve content production efficiency and user experience.

3.2.2 intelligent driven digital human: The digital human driven by a computer system to automatically complete a series of actions by technical means, including text-driven, audio-driven and video-driven.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

2D	two-Dimensional
3D	three-Dimensional
ASR	Automatic Speech Recognition
NLP	Natural Language Processing
TTS	Text To Speech

5 Conventions

This Recommendation uses the following conventions:

- The phrase "is required to" indicates a requirement that must be strictly followed and from which no deviation is permitted if conformity to this Recommendation is to be claimed.
- The phrase "is recommended" indicates a requirement that is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformity.

6 Framework for digital human application system

This Recommendation divides the digital human system into six parts: image; speech; animation; interactive processing; multimodal input; and multimodal output. The first four parts are closely related to digital human characters themselves, and the last two are the attributes that the digital human needs to have for presentation or interaction with multimodal input and output devices, as shown in Figure 1.