

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.745

(07/2016)

SERIES F: NON-TELEPHONE TELECOMMUNICATION
SERVICES

Multimedia services

Functional requirements for network-based speech-to-speech translation services

Recommendation ITU-T F.745



International
Telecommunication
Union
Documents

Document Sharing Hub

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 telemessage 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
HUMAN FACTORS	F.900–F.999

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

Recommendation ITU-T F.745

Functional requirements for network-based speech-to-speech translation services

Summary

Recommendation ITU-T F.745 specifies a high level functional model, a service description and requirements for speech-to-speech translation (S2ST) accomplished by connecting distributed S2ST modules all over the world through a network. To extend this network-based S2ST to other modalities, such as sign language, the modality conversion markup language (MCML) needs to have an expandable structure. The scope of this Recommendation is limited to the application protocol and the services using the network-based S2ST.

This revision includes additional information to clarify that Recommendation ITU-T F.745 could be applicable to both face-to-face communication and remote communication.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T F.745	2010-10-14	16	11.1002/1000/10982
2.0	ITU-T F.745	2016-07-14	16	11.1002/1000/12897

Keywords

Automatic speech recognition (ASR), machine translation (MT), modality conversion markup language (MCML), speech-to-speech translation (S2ST), text-to-speech synthesis (TTS).

* 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.....	2
4 Abbreviations and acronyms	3
5 Conventions	3
6 High-level functional model and generic service description	4
6.1 System overview	4
6.2 Functional model of modality conversion (MC) through communication between modality conversion protocol (MCP) clients and servers	4
6.3 Service description	7
7 Requirements	7
7.1 User input requirements	7
7.2 Network requirements	8
7.3 User device requirements	8
7.4 Modality conversion (MC) client requirements	8
7.5 Modality conversion (MC) server requirements	8
7.6 Quality requirements	9
7.7 Security and privacy requirements	9
7.8 Codec requirements	9
Appendix I – Service description in applications	10
I.1 Shared speech-to-speech translation (S2ST) client of two-party communication	10
I.2 Personal speech-to-speech translation (S2ST) client communication.....	10
I.3 Cross-modality communication.....	10
Bibliography.....	12

Introduction

The fact that the world has many different languages is one of the barriers to mutual understanding. The more directly people who speak different languages can communicate without language boundaries, the more mutual understanding can be accelerated and the closer human relationships can be constructed all over the world. To achieve such communication between humans, speech-to-speech translation (S2ST) technologies can be used. S2ST is a technology that recognizes the speech in one language, translates the recognized speech into another language, and then synthesizes the translation into speech. The leveraging of S2ST technologies in a pragmatic manner, which has long been one of mankind's dreams, may have a significant impact on tourism, social services, safety, and security by removing language barriers, and may ultimately influence language education. To construct S2ST systems, automatic speech recognition (ASR), machine translation (MT) and text-to-speech synthesis (TTS) must be built for source and target languages by collecting speech and language data, such as audio data, its manual transcriptions, pronunciation lexica for each word, parallel corpora for translation and so on. It is very difficult for individual organizations to build S2ST systems covering all topics and languages. However, by interconnecting ASR, MT and TTS modules developed by separate organizations and distributed globally through a network, one can create S2ST systems that break the world's language barriers.

This Recommendation defines the service description and the requirements for network-based S2ST technologies consisting of various distributed modules connected together in a network.