INTERNATIONAL STANDARD

ISO/IEC 20382-2

First edition
2017-10

Information technology — User interface — Face-to-face speech translation —

Part 2: System architecture and functional components

Technologies de l'information — Interface utilisateur — Face-à-face discours traduction —

Partie 2: Architecture du système et des composants fonctionnels

Reference number
ISO/IEC 20382-2:2017(E)

© ISO/IEC 2017



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2017, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword ... **iv**

Introduction ... **v**

1 Scope ... **1**

2 Normative references ... **1**

3 Terms, definitions and abbreviated terms ... **1**
3.1 Terms and definitions ... 1
3.2 Abbreviated terms ... 1

4 Overview of face-to-face speech translation ... **1**
4.1 General ... 1
4.2 Functional components of F2F speech translation ... 2

5 Functional requirements ... **2**
5.1 General requirement ... 2
5.2 Speech recognition requirements ... 3
5.3 Language translation requirements ... 3
5.4 Speech synthesizer requirements ... 3

6 System architectures of F2F speech translation ... **4**
6.1 General ... 4
6.2 Two persons with embedded F2F speech translation devices ... 4
6.3 Two persons with remote speech translation functions ... 6
6.4 Mixture of 6.2 and 6.3 ... 7
6.5 Adding one more speaker to F2F speech translation conversation ... 9
6.6 Two person with only one fixed F2F speech translation device ... 10

Annex A (informative) **History of F2F speech translation** ... **13**

Annex B (informative) **An example scenario of F2F speech translation protocol** ... **18**

Bibliography ... **19**

© ISO/IEC 2017 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form a specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organizations to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 35, *User interfaces*.

A list of all parts in the ISO/IEC 20382- series can be found on the ISO website.

© ISO/IEC 2017 – All rights reserved

Introduction

It is important to consider people with special requirements to ensure that they can gain the same benefits from ICT. One of those special requirements is to help people to avoid language barriers in global environments. Automatic speech translation systems have existed for a long time, but they have functional limitations as well as technical ones with regard to usability and accessibility. Annex A shows a history of face-to-face speech translation.

One reason for these limitations is the diversity of the languages currently used. It is difficult to support many languages by one or several speech translation systems. A flexible and interoperable standardized framework is needed to work with all different languages utilizing many speech translation systems already developed in many countries. Other considerations to make a natural and usable speech translation service possible include applying users' characteristics within the system, such as emotion, speech style, gender type and other attributes. To reflect those characteristics in the output speech translation, a standardized user interface is required to reflect the input and output data and transfer them to the user's device.

This document aims to enable face-to-face speech translation among people with different languages. The three technologies, i.e., speech recognition, language translation, and speech synthesis technologies, are mature enough to build a speech translation function. There are many face-to-face speech translation devices and/or services using mobile devices. However, the user needs to learn how to use the service and needs to use both hands to control the speech translation system. If the user wishes to use only one hand, which is usually the case, he or she cannot use the current speech translation systems and/or services. To overcome this usability issue, this document suggests a method that exactly follows the conversation among people with the same language. The method in this document is hands-free, and does not require any pre-training. In this sense, this method is the ultimate user interface of face-to-face speech translation and will open a world without language barriers.

© ISO/IEC 2017 – All rights reserved

Information technology — User interface — Face-to-face speech translation —

Part 2:
System architecture and functional components

1 Scope

This document specifies the functional components of face-to-face speech translation designed to interoperate among multiple translation systems with different languages. It also specifies the speech translation features, general requirements and functionality, thus providing a framework to support a convenient speech translation service in face-to-face situations. This document is applicable to speech translation devices, servers and communication protocols among speech translation servers and clients in a high-level approach. This document also defines various system architectures in different environments. This document is not applicable to defining speech recognition engines, language translation engines and speech synthesis engines.

2 Normative references

There are no normative references in this document.

3 Terms, definitions and abbreviated terms

3.1 Terms and definitions

No terms and definitions are listed in this document.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.2 Abbreviated terms

Utf-8 Unicode standard defined in IETF RFC 2279 (1998), UTF-8, a transformation format of ISO/IEC 10646

4 Overview of face-to-face speech translation

4.1 General

A face-to-face (F2F) speech translation system enables users of different languages in a face-to-face situation to communicate with each other in spoken languages by providing machine-generated translation results. A face-to-face speech translation system between a speaker and a listener shall have a speech recognition module, language translation module and a speech synthesizer (TTS: text to speech) as shown in Figure 1.

© ISO/IEC 2017 – All rights reserved