



**International
Standard**

ISO/IEC 23773-2

**Information technology —
User interfaces for automatic
simultaneous interpretation
systems —**

**Part 2:
Requirements and functional
description**

*Technologies de l'information — Interfaces utilisateur pour les
systèmes d'interprétation simultanée automatique —*

Partie 2: Exigences et description fonctionnelle

**First edition
2024-07**

IECNORM.COM : Click to view the full PDF of ISO/IEC 23773-2:2024



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2024

All rights reserved. Unless otherwise specified, or required in the context of its implementation, 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
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Abbreviated terms	1
5 Functional requirements	1
5.1 User interface (UI) requirements	1
5.1.1 General	1
5.1.2 Required	2
5.1.3 Optional	2
5.2 User device requirements	2
5.2.1 General	2
5.2.2 Required	2
5.2.3 Optional	2
5.3 Accessibility requirements	2
5.3.1 General	2
5.3.2 Required	2
5.3.3 Optional	3
6 Functional components of UI of automatic simultaneous interpretation systems	3
6.1 General	3
6.2 UI functional component	3
6.2.1 General	3
6.2.2 Setup function of the initial interpretation environments	3
6.2.3 Input and output function of the interpretation system	3
6.2.4 Correction/selection function of speech recognition/interpretation results	4
6.3 Accessibility functional component	4
6.3.1 General	4
6.3.2 Multi-modal input functions	4
6.3.3 Multi-modal output functions	4
6.3.4 Speech input and output functions	4
6.3.5 Accessibility functions for the application UI	4
6.3.6 Simple language or controlled language function	4
6.4 Simultaneous interpretation application functional component	5
Bibliography	6

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the 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 organization 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.

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 document 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 or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and <https://patents.iec.ch>. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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

For an explanation of 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 www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

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

A list of all parts in the ISO/IEC 23773 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

Communication between users of different languages is a global trend that is increasing. Real-time, automatic simultaneous interpretation is needed for different applications such as video calls, live lecture translation and wearable translation devices. Market demands for real-time automatic simultaneous interpretation of free-style continuous utterances in the travel sector, global event management, phone calls, lectures or meetings are also increasing. A standardized user interface (UI) for automatic simultaneous interpretation systems fulfils these different needs for communication.

The ISO/IEC 23773 series consists of the following parts:

ISO/IEC 23773-1 provides a general description of automatic simultaneous interpretation systems designed to interoperate among different natural languages for spontaneous speech and texts.

ISO/IEC 23773-2 (this document) provides the requirements and functional components for the UI of automatic simultaneous interpretation systems.

ISO/IEC 23773-3 provides a reference architecture for automatic simultaneous interpretation systems including functional modules and communication interfaces in a high-level approach.

IECNORM.COM : Click to view the full PDF of ISO/IEC 23773-2:2024

IECNORM.COM : Click to view the full PDF of ISO/IEC 23773-2:2024

Information technology — User interfaces for automatic simultaneous interpretation systems —

Part 2: Requirements and functional description

1 Scope

This document provides user interface requirements and functional description of real-time automatic simultaneous interpretation systems designed to interoperate among different natural languages for spontaneous speech.

While traditional speech to speech translation addresses the functional equivalent of consecutive interpretation, this document focuses on the functional equivalent of simultaneous interpretation.

This document does not cover sign language interpretation.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 23773-1, *Information technology — User interfaces for automatic simultaneous interpretation system — Part 1: General*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 23773-1 apply.

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

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <https://www.electropedia.org/>

4 Abbreviated terms

TTS text to speech

UI user interface

5 Functional requirements

5.1 User interface (UI) requirements

5.1.1 General

[Subclause 5.1](#) provides requirements regarding UI used for automatic simultaneous interpretation.

5.1.2 Required

- The interpretation system shall process the user's input in source language in text or speech form.
- The interpretation system shall present as the output the translation results in a target language in text or spoken form.
- The interpretation system shall provide an automatic simultaneous interpretation function for the continuous input.

5.1.3 Optional

- The interpretation system should allow the users to have a session with multiple users.
- The interpretation system should allow the users to have additional participants after the session starts.
- The interpretation system should allow the users to choose from different source languages for the input.
- The interpretation system should allow the users to choose from different target languages for the output.
- The interpretation system should allow the users to set up the preferences.
- The interpretation system should provide the translation results within 3 sec after the user's utterance finishes.
- The interpretation system should provide the translation results in speech reflecting the speaker's characteristics such as gender, pitch and power.
- The interpretation system should provide the translation results in speech reflecting the emotion of the speaker.

5.2 User device requirements

5.2.1 General

[Subclause 5.2](#) provides requirements regarding user devices used for automatic simultaneous interpretation.

5.2.2 Required

- The device shall support the input and output function of the automatic simultaneous interpretation.

5.2.3 Optional

- The device should provide a function to show to the user the confirmation signal to participate in the current translation session.
- The translation devices should be operated easily.

NOTE If the time to start operating a specific function takes more than 3 min, it does not satisfy this requirement.

5.3 Accessibility requirements

5.3.1 General

[Subclause 5.3](#) provides requirements regarding accessibility functions for users.

5.3.2 Required

- The input from the user shall be possible in a text form as well as a spoken form.
- The translation results shall be provided in a text form as well as a spoken form.

5.3.3 Optional

- The UI should provide multi-modal interface such as text, touch and gesture, to manage the interpretation application.
- The UI should provide the accessibility related user preference set-up functions.
- The translated results should also be provided in a controlled language.

NOTE Controlled languages are defined in ISO/TS 24620-1.

6 Functional components of UI of automatic simultaneous interpretation systems

6.1 General

The UI of the automatic simultaneous interpretation system consists of the following functional components:

- UI functional component;
- accessibility functional component;
- simultaneous interpretation application functional component.

The simultaneous interpretation application functional component plays a role in the interface between the UI and the simultaneous interpretation system. The simultaneous interpretation system is described in detail in ISO/IEC 23773-3. Its sub-modules include continuous speech recognition module, interpretation unit extraction module, real-time interpretation module, incremental knowledge learning module and presentation of translation results module.

6.2 UI functional component

6.2.1 General

The UI functional component has the following functions.

6.2.2 Setup function of the initial interpretation environments

- Start the simultaneous interpretation session;
- Select the source language and target language;
- Select input and output presentation option, e.g. speech, text only, mixed.

6.2.3 Input and output function of the interpretation system

The interpretation application component provides the interface for input and output of the automatic simultaneous interpretation system.

- Input function: The user's continuous speech input is sent through the application component to the continuous speech recognition module to be recognized. The text input will be directly sent to the interpretation unit extraction module.
- Output function: The translation results are sent to the presentation module either as speech signal or a text output and then to the interpretation application component to be presented to the user.

6.2.4 Correction/selection function of speech recognition/interpretation results

When the user finds errors in the recognized speech or interpretation results, they can correct the sentence by the following methods:

- keyboard correction;
- gesture correction;
- repeating the sentence with the error;
- selection of the correct sentence among several options.

In addition, the recognized sentence can be presented in as multiple results, each with its associated reliability, allowing the user to select the correct sentence. This error-correction function enables the system to learn from speech recognition and interpretation errors, facilitating the improvement of the system through incremental knowledge learning. This correction/selection function is meant for purpose of improvement of the system and testing/evaluation because it is not applicable to the real-time automatic simultaneous interpretation situation.

6.3 Accessibility functional component

6.3.1 General

An accessibility functional block provides the following functions for people with disabilities and specific needs.

6.3.2 Multi-modal input functions

The multi-modal input functions include text/touch/gesture input for persons who are hard of hearing who will use, for example, text to set up the interpretation application according to their preferences. When they want to use the translation function, they can also input their speech source language using a text keyboard instead of speech interface that will be translated through the application. The translated results can be displayed in text instead of speech so that the users can see the sentences in the target language. In the same way, touch and gesture can be used when the users select options for their translation input and output.

6.3.3 Multi-modal output functions

The multi-modal output functions include text/touch/gesture output for persons with hearing impairments. Persons with hearing disabilities will receive the translation results as text instead of speech so that they can read and understand what the other party said in a foreign language. The translation result can also be presented using sign language in the screen.

6.3.4 Speech input and output functions

The speech input and output functions are also parts of a multi-modal UI. Other than the intrinsic speech input and output of the translation system, a speech-enabled session setup function is provided for persons with visual impairments and persons with physical impairments. For example, when the user wants to set the source language as “English”, they can say “English” instead of typing it into the translation device

6.3.5 Accessibility functions for the application UI

The accessibility functions for the application UI include colour change, font enlargements for people with visual impairments and older people. For example, the older people can choose a bigger font when they use the translation device.

6.3.6 Simple language or controlled language function

The simple language or controlled language function (defined in ISO/TS 24620-1) provides simple output sentences which are easy to understand for persons with cognitive disabilities. For example, a simple

sentence with just one clause is easier to understand than that of multiple clauses. Vocabulary can be controlled for the easier understanding.

6.4 Simultaneous interpretation application functional component

The simultaneous interpretation application functional component is on the top layer of the interpretation system that provides the interface between the interpretation service and the system components. This component provides the functionality of sending the input to the interpretation system and receiving the translation results back as speech through text to speech (TTS) or text and providing them to the user. The simultaneous interpretation application module also provides the management of the application function for the initial user set-up of the session and selecting the preferences for a good user experience.

IECNORM.COM : Click to view the full PDF of ISO/IEC 23773-2:2024