



International
Standard

ISO/IEC 23773-1

First edition
2024-07

Information technology — User interfaces for automatic simultaneous interpretation systems —

Part 1: General

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

Partie 1: Généralités

IECNORM.COM : Click to view the full PDF of ISO/IEC 23773-1:2024

IECNORM.COM : Click to view the full PDF of ISO/IEC 23773-1:2024



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	2
5 General description of automatic simultaneous interpretation	2
6 Service flows and scenarios of automatic simultaneous interpretation systems	2
6.1 Service situations	2
6.2 Service flows and scenarios	4
7 General functions of automatic simultaneous interpretation systems	5
Bibliography	7

IECNORM.COM : Click to view the full PDF of ISO/IEC 23773-1:2024

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 (this document) 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 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 23773-1:2024

Information technology — User interfaces for automatic simultaneous interpretation systems —

Part 1: General

1 Scope

This document provides the general description of real-time automatic simultaneous interpretation systems designed to interoperate among different natural languages for spontaneous speech.

While traditional speech to speech machine 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

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions 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/>

3.1

incremental knowledge learning

knowledge accumulated through learning previous experiences in the context of translation

3.2

automatic simultaneous interpretation

automatic interpretation where the translation is performed continuously while the speaker speaks without waiting for the translation to finish sentence by sentence

Note 1 to entry: While interpretation deals with spoken language in real time, translation focuses on written content.

Note 2 to entry: Input and output are either speech or text, or both.

3.3

interpretation unit

user's utterance serving as the target for interpretation in simultaneous interpretation, facilitating a seamless and continuous translation process

4 Abbreviated terms

DB	database
TTS	text to speech
UI	user interface

5 General description of automatic simultaneous interpretation

An automatic simultaneous interpretation system helps people, for whom the language becomes a communication barrier, to communicate with each other through machine translation in a real-time situation. In an automatic simultaneous interpretation system, real-time interpretation for free-style continuous utterances is performed for travel purposes, global events, phone calls, lectures or meetings. The translation target is on each interpretation unit and incremental knowledge translation technology which applies. An automatic simultaneous interpretation system includes devices for processing input speech signals and output translation results. For different types of services, speech recognition and speech synthesis functions can be performed locally or remotely. The machine translation function resides in the translation servers, which can be performed either locally or remotely.

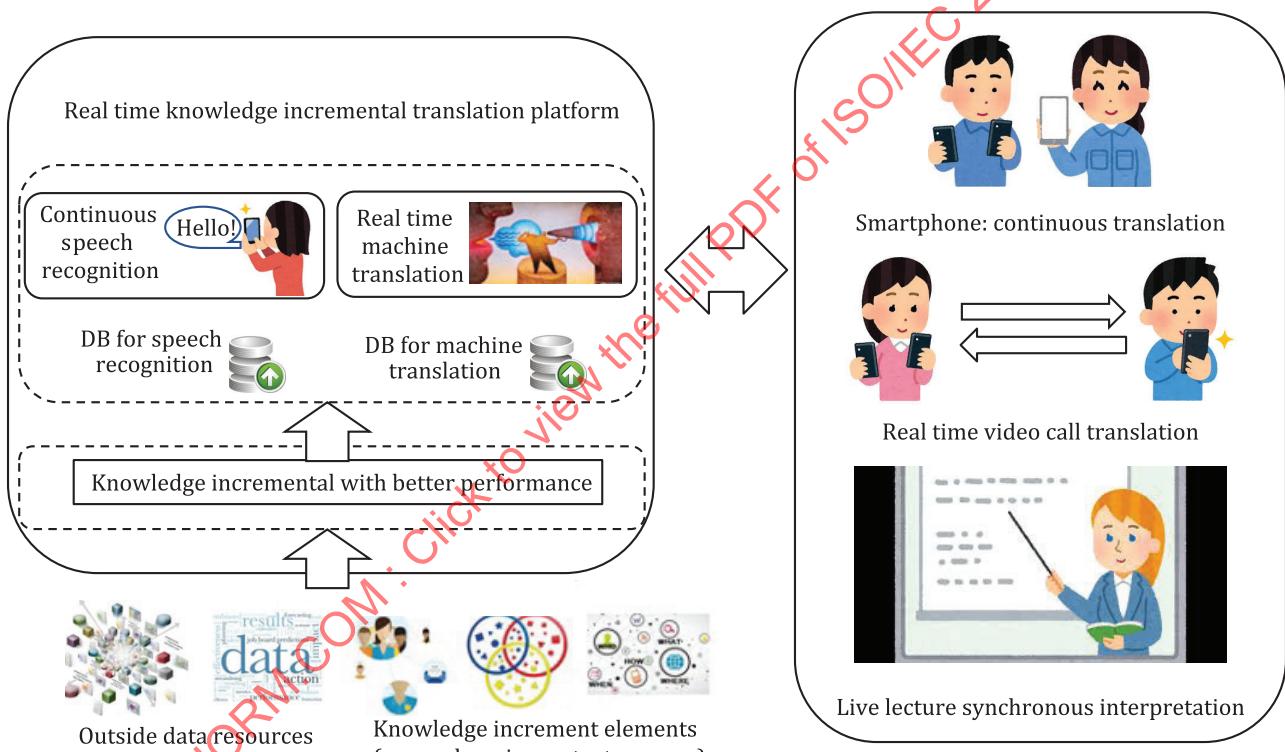


Figure 1 — Automatic simultaneous interpretation system and use cases

6 Service flows and scenarios of automatic simultaneous interpretation systems

6.1 Service situations

The user interface (UI) of automatic simultaneous interpretation systems can be classified according to the different interpretation situations described in this document. The characteristics of the different

situations are reflected in the interpretation services including UIs. The following use cases of the automatic simultaneous interpretation system in different situations are illustrated in [Figure 2](#) to [Figure 4](#):

- Interpretation for real-time public speech/live lecture ([Figures 2](#) and [3](#)): the speaker is usually one person and the interpretation service target is the audience, where people are using different languages. How many target languages are provided to the audience depends on the system.
- Interpretation for video talk: the participants are usually one-on-one and both people speak and need the simultaneous interpretation bi-directionally.
- Interpretation for face-to-face/on-line meeting: multiple people participate in the meeting and each person needs the simultaneous interpretation bi-directionally.
- Interpretation for phone calls: this is similar to the video talk except that there is no video involved.
- Interpretation for chat-bots/chatting: this is similar to an online meeting except that only text is used in most cases.



Figure 2 — Simultaneous interpretation of lectures in video clips

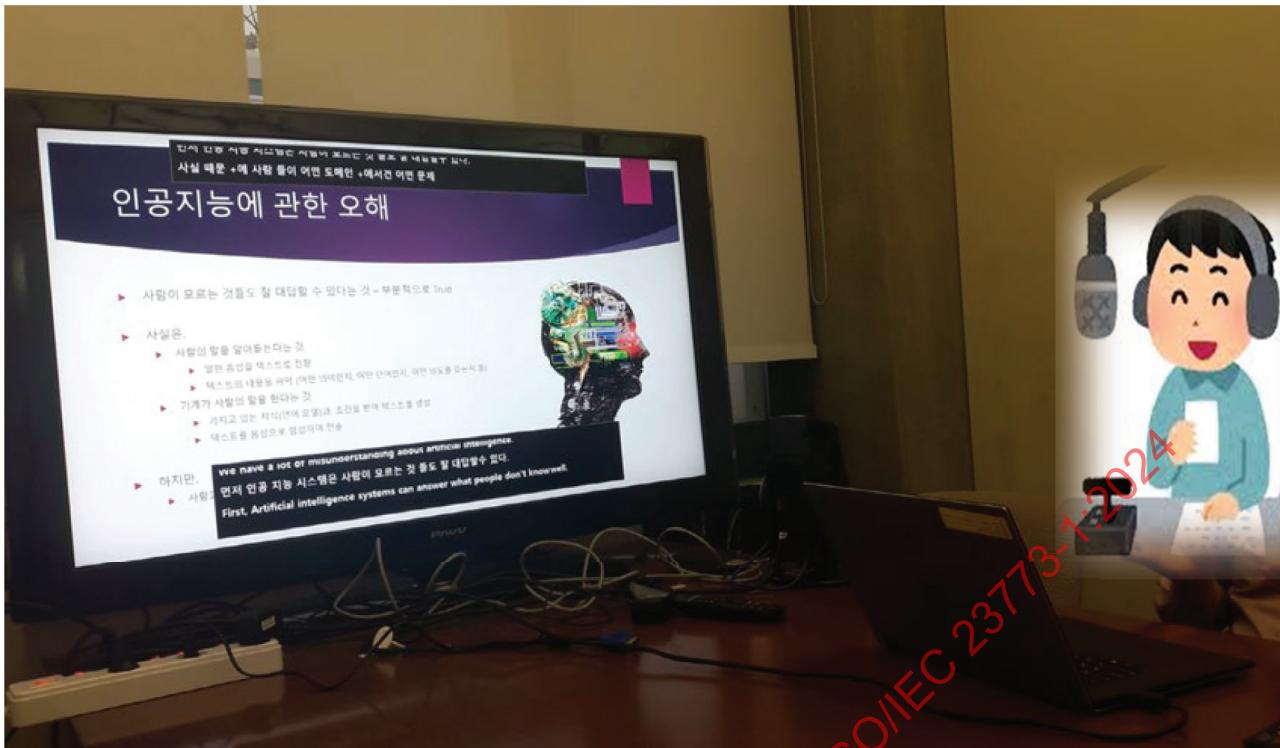


Figure 3 — Simultaneous interpretation of live lectures



After watching this, your brain will not be the same.

And why do some of us learn things more easily than others?

그리고 왜 우리의 일부가 다른 사람 보다 더 쉽게 배웁니까?

ETRI Multi-lingual simultaneous translation POC. Version Alpha x

Speech recognition results

And why do some of us learn things more easily than others?

Start translation

Stop translation

Sentence separation module output

So how do we learn?
And why do some of us learn things more easily than others?

Translation language setup

Korean → English
English → Korean
Korean → Chinese
Chinese → Korean
Korean → Spanish
Korean → French

Translation results

그럼 어떻게 배우나요?
그리고 왜 우리의 일부가 다른 사람 보다 더 쉽게 것을 배웁니까?

Figure 4 — Simultaneous interpretation of real-time video

6.2 Service flows and scenarios

The service flows and scenarios can be described differently depending on the interpretation situations which are introduced in 6.1. The following steps are typical real-time simultaneous interpretation service processes in the automatic simultaneous interpretation system. The real-time public speaking situation as shown in Figure 4 is assumed in this scenario.

- After a session starts, user A starts the real time speech in language A. The speech signal is transmitted continuously to the interpretation system.

- b) The speech is recognized by a continuous speech recognition module in the system. Then, the real-time interpretation operation is requested to the machine translation server.
- c) The translation is performed by the translation server and the translation result, which is in language B and is sent back in a text form.
- d) The translation result, which is in text form, is transformed to a speech signal by a speech synthesizer, or text to speech (TTS) along with the text in language B and continuously sent to the device of the interpretation service user for presentation as output speech or text, depending on the situation.

7 General functions of automatic simultaneous interpretation systems

The automatic simultaneous interpretation system consists of multiple functional components which are presented in [Figure 5](#) and described in the following list. A more detailed description of the functional components, their sub-modules and interfaces among them are defined in ISO/IEC 23773-2 and ISO/IEC 23773-3.

- Simultaneous interpretation application: The simultaneous interpretation application functional component is the top layer of the interpretation system that provides the interface between the interpretation service and the system components. UI functions such as input function and output function are provided in this functional component.
- Continuous speech recognition: Speech recognition is performed continuously on the speech units as sentence units or interpretation units from vocalized speech that is input in real time posed by the speech recognition engine that uses an acoustic model and a language model.
- Interpretation unit extraction: The real-time interpretation unit extraction functional component forms one or more of the speech units into an interpretation unit.
- Real-time simultaneous interpretation: The user speech is continuously translated into a target language based on the interpretation unit which is formed by the real-time interpretation unit extraction module to produce natural translation results without stopping.
- Incremental knowledge learning: Knowledge required for the speech recognition and translation is acquired from different knowledge sources, e.g. speech data, user log data, user and domain data and on-line data. The knowledge is incrementally learned and structured into a speech/translation knowledge database (DB) and a user/domain knowledge DB that will be accessed and used by the system processes.
- Presentation of translation results: The functional component of presentation of translation results processes the output of the interpretation system and manages the presentation control functions. The translation results are presented to the users in different output formats, such as text, speech or gestures, for different devices depending on the service types the interpretation system provides.