ISO/IEC~~-~~ PRF 23773-3~~:2024(E)~~

~~Date: 2024-0104~~

ISO~~-~~/IEC JTC~~-~~ 1/SC~~-~~ 35~~/WG 5~~

Secretariat:~~-~~ AFNOR

Date: 2024-04-30

Information technology~~-~~ — User interfaces for automatic simultaneous interpretation system~~—-~~ —

Part 3: System architecture

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

Partie 3: Architecture du système

FDIS stage

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
~~Fax: +41 22 749 09 47~~
~~Email~~E-mail: copyright@iso.org
Website: ~~www.iso.org~~www.iso.org

Published in Switzerland

Contents

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~~www.iso.org/directives or www.iec.ch/members_experts/refdocs).

Field Code Changed

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.~~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.~~www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

Field Code Changed

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~~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, as well as for phone calls, lectures or meetings, are also increasing. A standardized user interface (UI) for automatic simultaneous interpretation systems fulfils these different needs for communication.

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

While traditional speech-to-speech translation described in ISO/IEC 20382-1 and ISO/IEC 20382-2 addresses the functional equivalent of consecutive interpretation, this document focuses on the functional equivalent of simultaneous interpretation.

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

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

~~Title~~ Information technology ~~—~~ — User interfaces for automatic simultaneous interpretation system ~~—~~ —

Part 3: System architecture

1 Scope

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

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

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~~https://www.iso.org/obp

— ~~—~~IEC Electropedia: available at ~~https://www.electropedia.org/~~https://www.electropedia.org/

3.1
incremental knowledge learning
knowledge accumulated through learning from 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: Input is speech or text, or both, and output is speech or text, or both.

Note 2 to entry: While interpretation deals with spoken language in real time, translation focuses on written content.

3.3
interpretation unit
unit of the user's utterance which is the target for interpretation in the simultaneous interpretation in order to make the continuous translation

4 Abbreviated terms

ALT	alternative
APE	automatic post-editing
DB	database
DNN	deep neural network
KB	knowledge base
ML	machine learning
NMT	neural machine translation
POS	part of speech
RBMT	rule-based machine translation
RNN	recurrent neural network
LAS	listen and spell
LSTM	long short-term memory

5 Architecture of simultaneous interpretation system

5.1 General

The automatic simultaneous interpretation system consists of the following functional components which are presented in ~~Figure 1. Clause 5~~Figure 1. Clause 5 describes, in detail, the functional components, their sub-modules and interfaces among them. ~~Figure 2~~Figure 2 presents the flow chart of the automatic simultaneous interpretation to show how the functional components interact in the interpretation process.

- ~~—~~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.

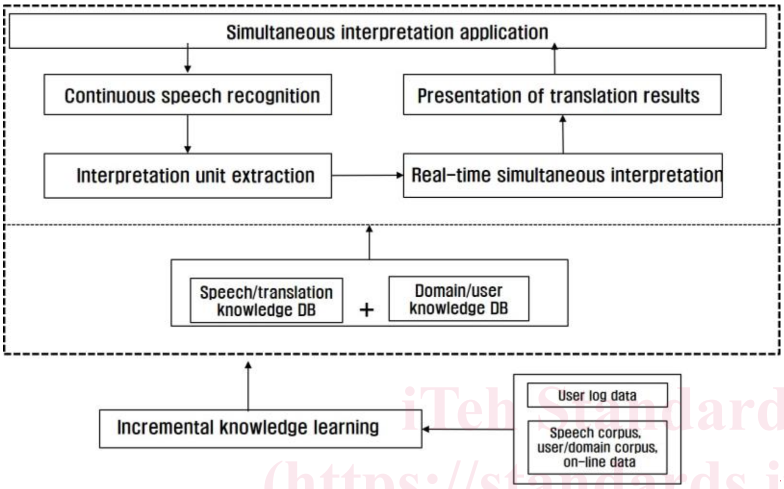
- ~~—~~Continuous speech recognition: Speech recognition is performed continuously on the speech units as sentence units or interpretation units from vocalized speech~~ that~~. The vocalized speech is input in real time posed by the speech recognition engine that uses an acoustic model and a language model. For more information of continuous speech recognition, see ISO/IEC 24661.

- ~~—~~Interpretation unit extraction: ~~a~~A 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 such as 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 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 including sign languages, for different devices depending on the service types that the interpretation system provides.


Simultaneous interpretation application
Continuous speech recognition
Presentation of translation results
Interpretation unit extraction
Real-time simultaneous interpretation
Speech/translation knowledge DB
+
Domain/user knowledge DB
Incremental knowledge learning
User log data
Speech corpus, user/domain corpus, on-line data

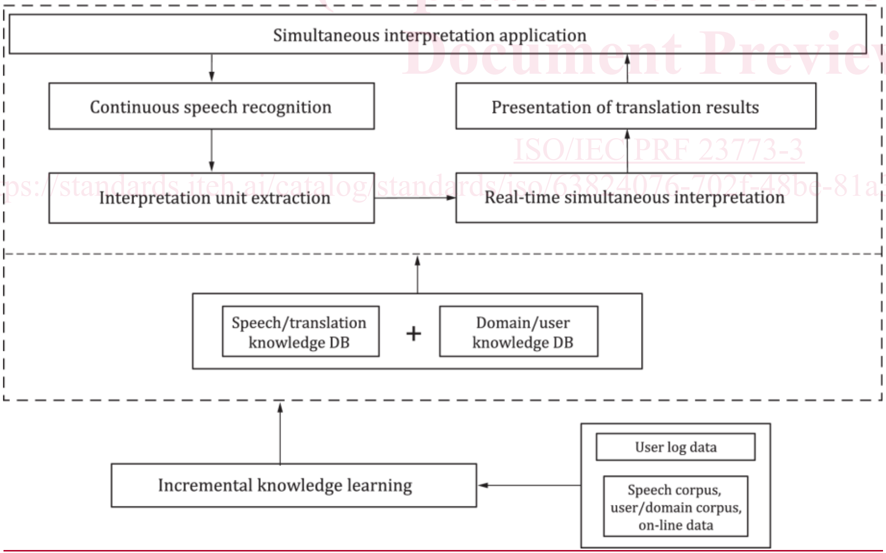

Simultaneous interpretation application
Continuous speech recognition
Presentation of translation results
Interpretation unit extraction
Real-time simultaneous interpretation
Speech/translation knowledge DB
+
Domain/user knowledge DB
Incremental knowledge learning
User log data
Speech corpus, user/domain corpus, on-line data

Figure 1 — Architecture of simultaneous interpretation system

ALT text for ~~Figure 1~~Figure 1

~~Figure 1~~Figure 1 consists of three big boxes. The first box (Box 1) is the biggest one and it is located at the top of the figure and the other two boxes (Box 2 and Box 3) are located horizontally under the first box. The three big boxes are connected ~~with~~by arrows. One arrow goes from Box 2 (Incremental knowledge learning) to Box 1. Another arrow goes from Box 3 to Box 2.

Box 1 consists of several boxes. The top box is labelled as “Simultaneous interpretation application”. There are four boxes under the top box, two on the right side and two on the left side. The first box under the left side of the top box is labelled as “Continuous speech recognition”. The second box under the left side of the top box is labelled as “Interpretation unit extraction”. The two boxes under the “Simultaneous interpretation application” are “Presentation of translation results” and then the “Real-time simultaneous interpretation” box under it. The arrow goes from the top box to the first box on the left and an arrow goes from that box to the second box on the left. A horizontal arrow goes from the second box on the left to the “Real-time simultaneous interpretation” box and another arrow goes up to the first box on the right. Finally, an arrow goes from the first box on the right to the top box. There are two boxes at the bottom of Box 1. They are “Speech/translation knowledge DB” box and “Domain/user knowledge DB”. The two boxes are connected to each other with a plus “+” mark. There is also an arrow from the plus “+” mark to the upper boxes.

Under the big Box 1, there is Box 2 ~~which is labelled as “~~(Incremental knowledge learning~~”~~) and it is connected by an arrow pointing to the big Box 1.

Finally, the last big box, Box 3 has two boxes vertically aligned: “User log data” box and “Speech corpus, User/Domain corpus, On-line data” box. An arrow points from Box 3 to Box 2~~.~~ (Incremental knowledge learning).