INTERNATIONAL STANDARD

ISO/IEC 24661

First edition
2023-05

Information technology — User interfaces — Full duplex speech interaction

Technologies de l'information — Interfaces utilisateur — Interaction vocale en duplex intégral

Reference number
ISO/IEC 24661:2023(E)

© ISO/IEC 2023



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2023

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

© ISO/IEC 2023 – All rights reserved

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Symbols and abbreviated terms ... **3**
5 Overview of FDX speech interaction UI ... **3**
5.1 Functional view ... 3
5.2 Main characteristics ... 4
5.2.1 General ... 4
5.2.2 Continuous ... 5
5.2.3 Natural ... 5
5.2.4 Adaptable ... 5
5.2.5 Initiative ... 5
5.2.6 Context-based ... 5
5.2.7 Knowledge-based ... 5
5.2.8 Model-based ... 5
6 Reference architecture of FDX speech interaction UI ... **5**
6.1 General ... 5
6.2 Interaction tasks ... 6
6.3 Functional components ... 7
6.3.1 General ... 7
6.3.2 Acoustic acquisition ... 7
6.3.3 Speech recognition ... 9
6.3.4 Conversation processing ... 10
6.3.5 Speech synthesis ... 12
6.4 Resources ... 12
6.4.1 Knowledge base ... 12
6.4.2 Data resources ... 13
6.5 Computing infrastructures ... 13
6.5.1 Cloud and edge computing ... 13
6.5.2 AI and ML systems ... 14
6.5.3 Network ... 14
7 Functional requirements and recommendations of FDX speech interaction UI ... **14**
7.1 General requirements and recommendations ... 14
7.2 Interaction task requirements and recommendations ... 15
7.3 Functional component requirements and recommendations ... 15
7.3.1 Acoustic acquisition requirements and recommendations ... 15
7.3.2 Speech recognition requirements and recommendations ... 15
7.3.3 Conversation processing requirements and recommendations ... 16
7.3.4 Speech synthesis requirements and recommendations ... 17
7.4 Resource requirements and recommendations ... 17
7.5 Computing infrastructures requirements and recommendations ... 17
8 Processes of FDX speech interaction UI ... **18**
8.1 General ... 18
8.2 Engineering process ... 18
8.3 Interaction process ... 19
9 Security and privacy considerations of FDX speech interaction UI ... **20**
Annex A (informative) **Example scenarios of FDX speech interaction** ... **21**
Bibliography ... **23**

© ISO/IEC 2023 – All rights reserved

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).

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) or the IEC list of patent declarations received (see https://patents.iec.ch).

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*.

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.

© ISO/IEC 2023 – All rights reserved

Introduction

Speech interaction user interface (UI) has been widely used for industrial applications and daily services. For example, it can be applied to automatic customer service in the telecommunication industry as a part of an interactive voice response system. From a communication point of view, a speech interaction UI can be recognized as a duplex-based system which enables bidirectional communication. In the early stages, speech interaction UIs for conventional dialogue systems were generally half duplex (HDX) based and were designed to be in a turn-oriented work mode. As the requirements of human-machine interaction have grown in complexity and diversity, the turn-oriented speech interaction UI has become unfit for a conversation between humans and machines.

Currently, full duplex (FDX) techniques are used in the speech interaction UI to support session-oriented conversations between humans and machines. The most significant differences between turn-oriented and session-oriented speech interactions are continuity and naturalness, which have made great progress in various applications of speech interaction UI, e.g. smart speaker, chatbot, intelligent assistant.

In recent years, a growing number of FDX speech interaction UIs have been studied and developed. This requires a common understanding of general models and specifications through standardization activities. In response to the standardization needs both from industry and academia, this document intends to provide a reference architecture, functional components and technical requirements of FDX speech interaction UI. For the benefit of system designers, developers, service providers and ultimate users, this document is composed of the following clauses:

— Clause 5 describes a functional view and general features of FDX speech interaction;

— Clause 6 provides a reference architecture and functional layers of FDX speech interaction UI;

— Clause 7 specifies the functional requirements regarding each functional layer;

— Clause 8 discusses the processes of FDX speech interaction UI;

— Clause 9 describes security and privacy considerations related to FDX speech interaction UI.

© ISO/IEC 2023 – All rights reserved

Information technology — User interfaces — Full duplex speech interaction

1 Scope

This document specifies user interfaces (UIs) designed for full duplex (FDX) speech interaction. It also specifies the FDX speech interaction model, features, functional components and requirements, thus providing a framework to support natural conversational interfaces between humans and machines. It also provides privacy considerations for applying FDX speech interaction.

This document is applicable to UIs for speech interaction and communication protocols for setting up a session-oriented FDX interaction between humans and machines.

This document does not define the speech interaction engines themselves or specify the details of specific engines, devices and approaches.

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
duplex
method of communication capable of transmitting data in both directions

[SOURCE: ISO 21007-1:2005, 2.18]

3.2
full duplex
FDX
method of communication capable of transmitting data in both directions at the same time

[SOURCE: ISO 21007-1:2005, 2.25]

3.3
functional unit
entity of hardware or software, or both, capable of accomplishing a specified purpose

Note 1 to entry: Functional units can be integrated as a system.

[SOURCE: ISO/IEC 2382:2015, 2123022, modified — Note 1 to entry has been changed and Note 2 and 3 to entry have been removed.]

© ISO/IEC 2023 – All rights reserved