INTERNATIONAL STANDARD

ISO/IEC 30122-2

First edition
2017-02

Information technology — User interfaces — Voice commands —

Part 2: Constructing and testing

Technologies de l'information — Interfaces utilisateurs — Commandes vocales —

Partie 2: Construire et tester

Reference number
ISO/IEC 30122-2:2017(E)

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 and definitions **1**
4 Conformity **2**
5 Requirements and recommendations **2**
5.1 General 2
5.2 Phonetics of phrase for a voice command 2
5.3 Design principle for a voice command 2
5.3.1 Phonetic description of words 2
5.3.2 Ease of pronunciation 2
5.3.3 Variation of pronunciation 2
5.4 Test of speech recognition engine 3
5.4.1 General 3
5.4.2 Subject 3
5.4.3 Samples of a voice command 3
5.4.4 Conditions of test 3
5.4.5 Procedure of test 3
5.4.6 Results 4
Bibliography **5**

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. 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 ISO 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 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 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.

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, Subcommittee SC 35, *User interfaces*.

A list of all parts in the ISO/IEC 30122 series can be found on the ISO website.

Introduction

Voice command is used for controlling ICT devices with the voice and in the language of the user. This technology is based on speech recognition, with some consideration for language tolerance (different accents or speech impairment while using a given language). It is also beneficial to the people who are operating the ICT device when/where they cannot use hands or fingers to operate it.

This document defines the principal standardized voice commands that will be commonly used in various ICT devices.

Information technology — User interfaces — Voice commands —

Part 2: Constructing and testing

1 Scope

This document provides the technical criterions and test methods of voice commands and its speech recognition engine.

The technical criterions include the phonetic requirements for spoken words or phrases that compose the voice command.

The test methods verify whether the voice command or speech recognition engine satisfies the required specifications.

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 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

3.1
word
smallest linguistic unit conveying a specific meaning

[SOURCE: ISO 5127:2001, 1.1.2.07, modified]

3.2
syllable
segment of speech which consists of a vowel with or without one or more accompanying consonants immediately preceding or following

[SOURCE: ISO 8253-3:2012, 3.8]

3.3
sentence
related group of word forms containing a predication, usually expressing a complete thought and forming the basic unit of discourse structure

[SOURCE: ISO 24615-1:2014, 3.15]

3.4
speech recognition
conversion, by a functional unit, of a speech signal to a representation of the content of the speech

[SOURCE: ISO/IEC 2382:2015, 2120735, modified]

4 Conformity

The voice command is conformant to this document if it meets all of the requirements of Clause 5.

5 Requirements and recommendations

5.1 General

Words or phrases for voice commands shall be recognized correctly. Therefore, the following conditions are considered.

— Words or phrases for voice commands are recognized easily.

— Acoustically similar words or phrases are not registered.

— Words or phrases that waver depending on the individual are not used.

5.2 Phonetics of phrase for a voice command

The voice command phrase should include plural words.

The words that are used in the phrase for a voice command should consist of plural syllables.

5.3 Design principle for a voice command

5.3.1 Phonetic description of words

Pronunciations of words used in a voice command should be described in an appropriate mark-up way.

EXAMPLE International Phonetic Alphabet (IPA), Speech Assessment Methods Phonetic Alphabet (SAMPA), etc.

NOTE A speech recognition engine does not always distinguish the difference between intonations or accents.

5.3.2 Ease of pronunciation

Words used in a voice command should be able to be pronounced easily by a native person, and ideally also by a person with language difficulties such as dyslalia, dysphemia, etc.

The length of phrase for a voice command should be able to be spoken in one breath.

NOTE Combination of words, such as a tongue twister, are difficult to pronounce.

5.3.3 Variation of pronunciation

If a word that has more than two pronunciations is used, all variations of pronunciations shall be registered to the voice command attribute (see ISO/IEC 30122-1).

5.4 Test of speech recognition engine

5.4.1 General

This subclause provides how to test the speech recognition engine to be used to recognize voice commands. This subclause provides only test and evaluation methods and does not provide quantitative criterions of specification.

5.4.2 Subject

At least 10 subjects shall participate in the test. Ideally, not less than 100 subjects should participate.

Subjects should be selected from diverse demographic groups such as the following:

a) age;

b) gender;

c) native language;

d) culture;

e) physical condition (person with or without disability);

f) geographic location.

Subjects can be classified into the following four groups. Subjects from all four groups should be included as participants for the test.

a) subjects who are using voice commands frequently;

b) subjects who have experience using voice commands;

c) subjects who have no experience using voice commands;

d) blind and visually impaired people could be associated for a different test for evaluation.

5.4.3 Samples of a voice command

Samples of a voice command to be used for the test should be selected from the following groups:

a) voice commands that are frequently used;

b) voice commands that are important for system specification;

c) voice commands that are estimated to be difficult for a speech recognition engine to recognize;

d) voice commands that are chosen randomly.

The number of samples of a voice command shall be at least 10. It should be as many as possible.

5.4.4 Conditions of test

The speech recognition engine should be tested in realistic conditions of use.

EXAMPLE A car navigation test is conducted in the same signal-to-noise ratio condition of driving.

5.4.5 Procedure of test

The subject speaks the sample of a voice command, and the spoken sample is directly provided or is recorded at once and will be later provided to the speech recognition engine.