INTERNATIONAL STANDARD

ISO
9241-154

First edition
2013-02-01

Ergonomics of human-system interaction —

Part 154:
Interactive voice response (IVR) applications

Ergonomie de l'interaction homme-système —

Partie 154: Applications de réponse vocale interactive (RVI)

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 9241-154:2013
https://standards.iteh.ai/catalog/standards/sist/137442a8-0ac0-4d12-b7de-
20576c34f602/iso-9241-154-2013

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 9241-154 was prepared by Technical Committee ISO/TC 159, *Ergonomics*, Subcommittee SC 4, *Ergonomics of human–system interaction*.

ISO/TS 9241 consists of the following parts, under the general title *Ergonomic requirements for office work with visual display terminals (VDTs)*:

— *Part 1: General introduction*

— *Part 2: Guidance on task requirements*

— *Part 5: Workstation layout and postural requirements*

— *Part 6: Guidance on the work environment*

— *Part 11: Guidance on usability*

— *Part 12: Presentation of information*

— *Part 13: User guidance*

— *Part 14: Menu dialogues*

— *Part 15: Command dialogues*

— *Part 16: Direct manipulation dialogues*

ISO 9241 also consists of the following parts, under the general title *Ergonomics of human-system interaction*:

— *Part 20: Accessibility guidelines for information/communication technology (ICT) equipment and services*

— *Part 100: Introduction to standards related to software ergonomics* [Technical Report]

— *Part 110: Dialogue principles*

— *Part 129: Guidance on software individualization*

— *Part 143: Forms*

— *Part 151: Guidance on World Wide Web user interfaces*

— *Part 154: Interactive voice response (IVR) applications*

— *Part 171: Guidance on software accessibility*

— *Part 210: Human-centred design for interactive systems*

— *Part 300: Introduction to electronic visual display requirements*

— *Part 302: Terminology for electronic visual displays*

— *Part 303: Requirements for electronic visual displays*

— *Part 304: User performance test methods for electronic visual displays*

— *Part 305: Optical laboratory test methods for electronic visual displays*

— *Part 306: Field assessment methods for electronic visual displays*

— *Part 307: Analysis and compliance test methods for electronic visual displays*

— *Part 308: Surface-conduction electron-emitter displays (SED)* [Technical Report]

— *Part 309: Organic light-emitting diode (OLED) displays* [Technical Report]

— *Part 310: Visibility, aesthetics and ergonomics of pixel defects* [Technical Report]

— *Part 331: Optical characteristics of autostereoscopic displays* [Technical Report]

— *Part 391: Requirements, analysis and compliance test methods for the reduction of photosensitive seizures*

— *Part 400: Principles and requirements for physical input devices*

— *Part 410: Design criteria for physical input devices*

— *Part 411: Evaluation methods for the design of physical input devices* [Technical Specification]

— *Part 420: Selection of physical input devices*

— *Part 910: Framework for tactile and haptic interaction*

— *Part 920: Guidance on tactile and haptic interactions*

The following parts are under preparation:

— *Part 940: Evaluation of tactile and haptic interactions*

User-interface elements, ergonomic requirements for the reduction of visual fatigue from stereoscopic images, and framework and guidance for gesture interactions are to form the subjects of future parts 161, 392 and 960.

Introduction

This part of ISO 9241 contains provisions specific to interactive voice response (IVR) systems, which may involve a combination of voice technologies, but are distinguished by the use of the telephone as the information transfer mechanism. These provisions assume no visual displays of information to the user beyond the labels on the telephone's keypad, with the notable exception of text telephones (TTYs), which have a visual feedback display.

Although it is extremely important that IVR user interface designers take into account the cultural and linguistic aspects of the user interface that impact the intended user population, these aspects are beyond the scope of this part of ISO 9241 and are not addressed in this part of ISO 9241 . Similarly, because automatic speech recognition (ASR) performs differently for different languages and the technology continues to improve, it is beyond the scope of this document to provide detailed provisions on ASR user interface design generally. Rather, this part of ISO 9241 focuses on the design of IVR dialogues and discusses only those ASR user interface design issues that impact dialogue design.

Many current IVR systems pose significant accessibility challenges to callers with disabilities. Some of the provisions in this part of ISO 9241 were developed specifically to accommodate callers with special needs, particularly those who are deaf or who have hearing impairments.

The provisions in this document are intended to be compatible with ISO/IEC 13714.

Interactive voice response (IVR) systems became a common means of delivering customer service in the late 1980s. These systems are designed to reduce or eliminate the need for human-in-the-loop customer support by automating many of the functions that human customer service representatives typically provide over the telephone with respect to processing of customer transactions. Thus, users (i.e. callers) can now engage in such activities as checking train schedules, ordering a book or reporting problems with their television cable service by interacting with an IVR system. In addition, IVRs often automate call-routing functions so that the caller can be connected with the right assistance to handle their specific request.

This part of ISO 9241 is concerned with the design of the human–IVR system dialogue and related topics. As shown in Figure 1, callers typically interact with the IVR system through one of two methods: speech or touchtone (DTMF) input via the telephone keypad. If an IVR system is speech-enabled, it employs an ASR engine that recognizes the speech input from the caller. If it is not speech-enabled, it typically recognizes only touchtone input from the telephone keypad or, sometimes, TTY input. Speech-enabled IVR systems are a relatively recent development and many systems are now designed to accept both touchtone and speech within a given dialogue with a caller. The IVR system responds via hardware and/or software that presents synthesized, digitized or recorded speech to the caller and that may also present non-speech audio. The fact that there is no assumed visual display of information to the caller in these applications poses a challenge to dialogue designers because of the burden placed on the caller to navigate the application and process and remember the relevant information without the aid of any visual display.
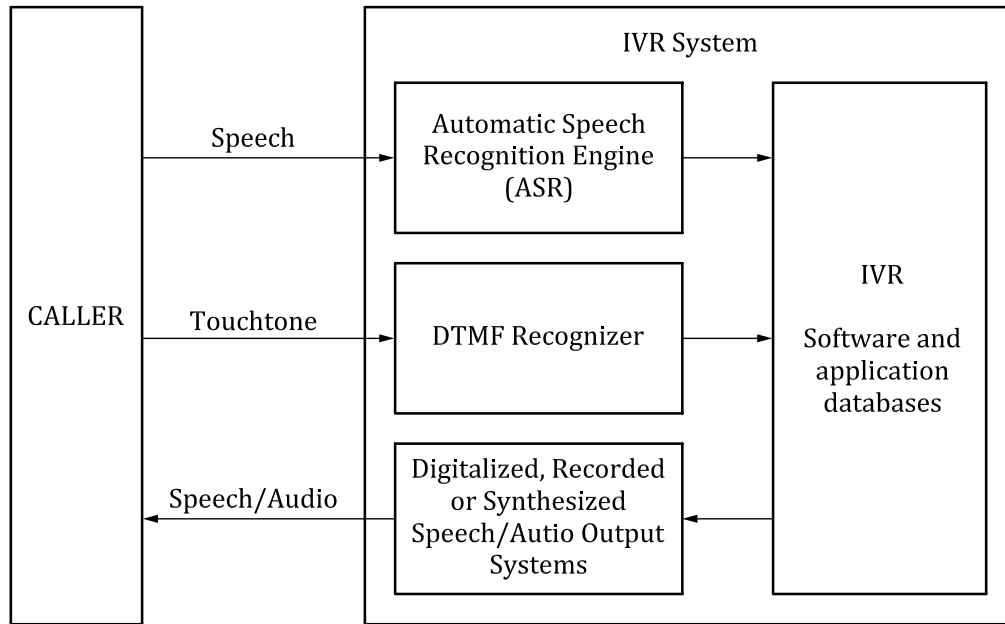
Figure 1 — IVR system representation

This part of ISO 9241 provides provisions for the design of IVR dialogues in speech-enabled and touchtone-based IVR systems. Therefore, its focus is on the interface between the caller and IVR software and application databases, which is mediated by the hardware and software that recognize speech and/or touchtones, and which present speech output to the caller. Both ASR systems and DTMF recognizers place constraints on the design of IVR dialogues and those constraints have been considered in developing the provisions of this part of ISO 9241. However, this document is not intended to address the design of ASR or DTMF user interfaces generally, for two reasons: first, the technology, particularly for ASR systems, is still evolving, and second, many ASR user interfaces are designed to include a visual display (e.g. ASR dictation applications), which IVR applications are not assumed to have.

Lastly, most of the provisions for speech-enabled dialogues are intended for use with grammar-based IVRs. Although some of the same principles apply to natural language systems (i.e. applications that use statistical language models), detailed design provisions are not included for these applications in this part of ISO 9241 because natural language understanding is implemented via a distinct technology and the use of natural language in speech-enabled IVRs is still evolving. There are also some aspects of applications design that are different for natural language dialogues, relative to grammar-based ones, in addition to the differences in dialogue design that relate to speech recognizers specifically.

Ergonomics of human-system interaction —

Part 154:
Interactive voice response (IVR) applications

1 Scope

This part of ISO 9241 gives guidance on, and requirements for, the user interface design of interactive voice response (IVR) applications. It covers both IVR systems that employ touchtone input and those using automated speech recognition (ASR) as the input mechanism. It is equally applicable to cases in which the caller or the IVR system itself (e.g. in some telemarketing applications) initiates the call.

This part of ISO 9241 is intended to be used together with ISO/IEC 13714.

NOTE   Its scope is thus more general than that of ISO/IEC 13714, which is specific to voice messaging systems.

2 Normative references

The following referenced documents are indispensable for the application 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 13714, *Information technology — Document processing and related communication — User interface to telephone-based services — Voice messaging applications*

ITU-T E 161, *Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1
announcement
message presented by the IVR to the caller, which informs but does not instruct the caller to act

Note 1 to entry: Some industry standards do not distinguish between announcements and prompts and consider all system-originated messages to be prompts.

Note 2 to entry: Prompts specifically instruct the caller except for subsequent input from the caller (see 3.33).

3.2
automatic speech recognition
automated speech recognition
ASR
conversion of spoken words to machine-readable input

3.3
barge-in
capability of an IVR system to accept input while a prompt (or an announcement) is being played

Note 1 to entry: The playback of speech ceases immediately and the system responds to the input of the caller.

Note 2 to entry: See *dial-through* (3.13) and *talk-through* (3.43). The term "barge-in" is synonymous with the term "dial-through" for systems employing touchtone input. It is also synonymous with "talk-through" for systems accepting speech input.

3.4
caller
user who calls for or is called by a service, gets connected to the IVR system, and interacts with it

Note 1 to entry: In an IVR system the caller is considered to be synonymous with the user.

3.5
coaching
instructions to the speakers of the prompts and announcements (also known as the voice talent) about desired subtleties in the prompt recordings

3.6
concatenated prompts
prompts or announcements constructed by stringing together several individual prompts or announcements

Note 1 to entry: Concatenated prompts are usually employed for the presentation of information that is dynamic or context-specific.

3.7
confidence level
commitment with which the speech recognition system returns a recognition result for a given input

Note 1 to entry: See *confidence score* (3.8).

Note 2 to entry: Confidence levels are often defined by ranges of confidence scores, which are usually classified as High, Medium and Low. Ranges are then used to trigger certain IVR responses, including acceptance, rejection or confirmation of caller speech.

3.8
confidence score
score assigned by the ASR system indicating its degree of confidence that it correctly recognized the caller's utterance

Note 1 to entry: For every caller input, the ASR system returns a list of potential candidate words or phrases with a numeric score representing the hypothesized probability of correctness.

3.9
continuous speech recognition
speech recognition that allows continuous input of words or phrases

Note 1 to entry: No minimum duration of silence is required at the beginning or end of words or phrases (other than the pauses generally occurring in natural speech patterns). Contrast with "Discrete speech recognition".

3.10
conversational repair
capability of an IVR system to resolve a conversational breakdown that occurs due to a speech or touchtone input error, by making use of subsequent input by the caller to determine the appropriate next step in the dialogue

EXAMPLE   The caller in a travel arrangements application, when asked his destination, says "Buffalo". The IVR system replies, "You want to go to Chicago, correct?" In response, the caller says, "No, Buffalo." If the system employs conversational repair, it would be able to determine that the caller has provided "corrective" input and would attempt to recognize that input and use it to move the dialogue along. Its response might then be. "Oh, you meant Buffalo. I'm sorry. When would you like to depart?" rather than asking the caller the same initial question again (i.e. "Where do you want to go?").

3.11
deletion error
instance of a recognition error where part of a speaker's utterance is incorrectly omitted in the speech recognizer's output

3.12
dial-ahead
capability of an IVR system to accept touchtone input before the system has requested it

Note 1 to entry: The touchtone input is then used by the system based on the order in which the input was received. This allows callers to provide input without having to listen to the associated input prompts.

Note 2 to entry: See *talk-ahead* (3.42).

3.13
dial-through
capability in a touchtone-based IVR system to accept caller input while a prompt (or an announcement) is being played

Note 1 to entry: See *talk-through* (3.43) and *barge-in* (3.3).

Note 2 to entry: In response to dial-through, the playback of speech ceases and the system responds to the key that was pressed.

3.14
digitized speech
digital recording of human speech

iTeh STANDARD PREVIEW

Note 1 to entry: Contrast with *synthesized speech* (3.40), *recorded speech* (3.35) and *text-to-speech* (3.45).

(standards.iteh.ai)

3.15
discourse marker
ISO 9241-154:2013
word, phrase or sound that is used as an indication to the caller that a new prompt or announcement is starting or that the caller is now expected to provide input
https://standards.iteh.ai/catalog/standards/sist/137442a8-0ac0-4d12-b7de-20576c34f602/iso-9241-154-2013

Note 1 to entry: Common discourse markers are "okay," "alright," and "now."

3.16
discrete speech recognition
speech recognition that requires a silence of some minimum duration at the beginning and at the end of the word or phrase to be recognized, to allow proper parsing of speech input

Note 1 to entry: Contrast with *continuous speech recognition* (3.9).

3.17
dual tone multiple frequency
DTMF
touchtones of the contemporary telephone keypad

3.18
dynamic grammar
grammar which is not predetermined and that is used for speech recognition

EXAMPLE   A caller is asked to speak a prescription number for refill and the grammar is composed of only that caller's prescription numbers, not all possible numeric combinations.

Note 1 to entry: It is usually created in real-time based on variable data.

**3.19
end-pointing**
process intended to detect the beginning and end of speech input

Note 1 to entry: In IVR the beginning of speech is typically referred to as onset and the ending of speech is typically referred to as offset.

**3.20
enrolment**
procedure by which callers provide training input for speaker-dependent speech recognition systems that require training before they can be used

**3.21
explicit confirmation**
method whereby the caller is prompted to confirm his or her input to an IVR system

Note 1 to entry: Contrast with implicit confirmation.

**3.22
grammar**
body of syntactic, and sometimes morphologic, rules defining all caller utterances that are covered by the speech recognition system

Note 1 to entry: The grammar specifies the input that is recognized by the application.

**3.23
implicit confirmation**
method of confirming the caller's input to an IVR system in which the IVR system responds to that input as if the input was correct

EXAMPLE 1   If a caller inputs "balance" to indicate that she wants the system to give her the balance in her bank account, the IVR's response of "Your balance is $452.19" implicitly indicates that it correctly recognized the request for "balance" information, as opposed to some other information about the account.

EXAMPLE 2   "Tomorrow." – IVR system: "At what time do you want to leave tomorrow?"

Note 1 to entry: In the case of implicit confirmation, the caller knows what input was received based on the subsequent action of the IVR system

Note 2 to entry: Implicit confirmations are a natural way of speeding up the dialogue. However, in the case of an error (disconfirmation), the mechanism of dialogue repair is not clear.

**3.24
insertion error**
instance of a recognition error where one or more words in the speech recognizer's output do not correspond with any word (or sequence of words) in the speaker's utterance

**3.25
interactive voice response
IVR**
software application that a caller interacts with over a telephone line and which presents pre-recorded and/or dynamically generated speech output and which can accept touchtone and/or speech input from the caller

**3.26
landmark**
<IVR> short phrase or a sound that acts as a heading or launch point for a portion of an IVR dialogue and which can be used to facilitate navigation of the caller within an IVR application

EXAMPLE   Short phrase: "Account Information", "Repair Services".

3.27
message
<IVR> information in an IVR system that is provided by the system, other callers, other subscribers, or system administrators to the caller

Note 1 to entry: Messages include both prompts (i.e. instructions for action) and announcements (i.e. no action required).

3.28
mixed initiative
IVR flow that contains both caller and system initiative situations

3.29
natural language understanding
NLU
technology used to recognize certain words and phrases from a caller utterance spoken as if talking with another human

Note 1 to entry: NLU does not actually understand the caller, but it is used with other technologies to derive the caller's meaning.

3.30
open-ended dialogue
dialogue that does not constrain the verbal responses from a caller

EXAMPLE   "What can I do for you?"

Note 1 to entry: This is typically used in conjunction with NLU.

3.31
persona
<IVR> set of personal, human characteristics conveyed by the application through the speakers of prompts and announcements (sometimes called the voice talent), word choices for prompts and other stylistic/aesthetic qualities of the IVR

Note 1 to entry: An IVR application persona differs from persona as used in other areas of human–computer interface design. An IVR application persona gives the caller an impression of the company or organization that is the focus of the application (e.g. a bank services centre, a retail company), as represented by the voice that conveys the prompts and announcements within the application. In other areas of human-computer interface design, a persona is a detailed description of a representative user that is used to guide application design.

3.32
priming
inclusion of example phrases or other prompt techniques to influence the caller's utterance style and length in speech-enabled IVR systems

EXAMPLE 1   An IVR banking application prompts caller word choices for menu items by using the specific words it wants the caller to say in its prompt: "You can pay bills, check your account balance, or make a deposit."

EXAMPLE 2   "Tell us what you would like to do. You can say 'pay bills', 'check account balance' or 'make a deposit." Here, the application tells the caller exactly what to say.

Note 1 to entry: This is typically used for open-ended dialogues.

3.33
prompt
system output requesting input from the caller

Note 1 to entry: Contrast with *announcement* (3.1).

Note 2 to entry: This term is often used generically to mean any message played by the IVR.