भारतीय मानक
Indian Standard

IS 16595 (Part 154) : 2018
ISO 9241-154 : 2013

मानव–प्रणाली बातचीत के श्रमदक्षता शास्त्र

भाग 154 इंटरएक्टिव वॉयस रिस्पांस (आइवीआर) अनुप्रयोग

Ergonomics of Human-System Interaction

Part 154 Interactive Voice Response (IVR) Applications

ICS 13.180; 35.180

भारतीय मानक ब्यूरो
BUREAU OF INDIAN STANDARDS
मानक भवन, 9 बहादुरशाह ज़फर मार्ग, नई दिल्ली–110002
MANAK BHAVAN, 9 BAHADUR SHAH ZAFAR MARG
NEW DELHI-110002
www.bis.gov.in www.standardsbis.in

January 2018 **Price Group 12**

Ergonomics Sectional Committee, PGD 15

NATIONAL FOREWORD

This Indian Standard (Part 154) which is identical with ISO 9241-154 : 2013 'Ergonomics of human-system interaction — Part 154: Interactive voice response (IVR) applications' issued by the International Organization for Standardization (ISO) was adopted by the Bureau of Indian Standards on the recommendation of the Ergonomics Sectional Committee and approval of the Production and General Engineering Division Council.

Besides Part 154, ISO has issued the following parts of ISO 9241 series so far:

Sl. No.	*ISO Standard*	*Title*
1.	ISO 9241-1 : 1997	Ergonomic requirements for office work with visual display terminals (VDTs) — Part 1: General introduction and ISO 9241-1 : 1997/Amd 1 : 2001
2.	ISO 9241-2 : 1992	Ergonomic requirements for office work with visual display terminals (VDTs) — Part 2: Guidance on task requirements
3.	ISO 9241-5 : 1998	Ergonomic requirements for office work with visual display terminals (VDTs) — Part 5: Workstation layout and postural requirements
4.	ISO 9241-6 : 1999	Ergonomic requirements for office work with visual display terminals (VDTs) — Part 6: Guidance on the work environment
5.	ISO 9241-11 : 1998	Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability
6.	ISO 9241-12 : 1998	Ergonomic requirements for office work with visual display terminals (VDTs) — Part 12: Presentation of information
7.	ISO 9241-13 : 1998	Ergonomic requirements for office work with visual display terminals (VDTs) — Part 13: User guidance
8.	ISO 9241-14 : 1997	Ergonomic requirements for office work with visual display terminals (VDTs) — Part 14: Menu dialogues
9.	ISO 9241-15 : 1997	Ergonomic requirements for office work with visual display terminals (VDTs) — Part 15: Command dialogues
10.	ISO 9241-16 : 1999	Ergonomic requirements for office work with visual display terminals (VDTs) — Part 16: Direct manipulation dialogues
11.	ISO 9241-20 : 2008	Ergonomics of human-system interaction — Part 20: Accessibility guidelines for information/communication technology (ICT) equipment and services
12.	ISO/TR 9241-100 : 2010	Ergonomics of human-system interaction — Part 100: Introduction to standards related to software ergonomics
13.	ISO 9241-110 : 2006	Ergonomics of human-system interaction — Part 110: Dialogue principles
14.	ISO 9241-129 : 2010	Ergonomics of human-system interaction — Part 129: Guidance on software individualization
15.	ISO 9241-143 : 2012	Ergonomics of human-system interaction — Part 143: Forms
16.	ISO 9241-151 : 2008	Ergonomics of human-system interaction — Part 151: Guidance on World Wide Web user interfaces
17.	ISO 9241-161 : 2016	Ergonomics of human-system interaction — Part 161: Guidance on visual user-interface elements
18.	ISO 9241-171 : 2008	Ergonomics of human-system interaction — Part 171: Guidance on software accessibility
19.	ISO 9241-210 : 2010	Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems

Sl. No.	*ISO Standard*	*Title*
20.	ISO 9241-300 : 2008	Ergonomics of human-system interaction — Part 300: Introduction to electronic visual display requirements
21.	ISO 9241-302 : 2008	Ergonomics of human-system interaction — Part 302: Terminology for electronic visual displays
22.	ISO 9241-303 : 2011	Ergonomics of human-system interaction — Part 303: Requirements for electronic visual displays
23.	ISO 9241-304 : 2008	Ergonomics of human-system interaction — Part 304: User performance test methods for electronic visual displays
24.	ISO 9241-305 : 2008	Ergonomics of human-system interaction — Part 305: Optical laboratory test methods for electronic visual displays
25.	ISO 9241-306 : 2008	Ergonomics of human-system interaction — Part 306: Field assessment methods for electronic visual displays
26.	ISO 9241-307 : 2008	Ergonomics of human-system interaction — Part 307: Analysis and compliance test methods for electronic visual displays
27.	ISO/TR 9241-308 : 2008	Ergonomics of human-system interaction — Part 308: Surface-conduction electron-emitter displays (SED)
28.	ISO/TR 9241-309 : 2008	Ergonomics of human-system interaction — Part 309: Organic light-emitting diode (OLED) displays
29.	ISO/TR 9241-310 : 2010	Ergonomics of human-system interaction — Part 310: Visibility, aesthetics and ergonomics of pixel defects
30.	ISO/TR 9241-331 : 2012	Ergonomics of human-system interaction — Part 331: Optical characteristics of autosterescopic displays
31.	ISO 9241-391 : 2016	Ergonomics of human-system interaction — Part 391: Requirements, analysis and compliance test methods for the reduction of photosensitive seizures
32.	ISO 9241-392 : 2015	Ergonomics of human-system interaction — Part 392: Ergonomic recommendations for the reduction of visual fatigue from stereoscopic images
33.	ISO 9241-400 : 2007	Ergonomics of human-system interaction — Part 400: Principles and requirements for physical input devices
34.	ISO 9241-410 : 2008	Ergonomics of human-system interaction — Part 410: Design criteria for physical input devices & ISO 9241-410 : 2008/Amd 1 : 2012
35.	ISO/TS 9241-411 : 2012	Ergonomics of human-system interaction — Part 411: Evaluation methods for the design of physical input devices
36.	ISO 9241-420 : 2011	Ergonomics of human-system interaction — Part 420: Selection of physical input devices
37.	ISO 9241-910 : 2011	Ergonomics of human-system interaction — Part 910: Framework for tactile and haptic interaction
38.	ISO 9241-920 : 2009	Ergonomics of human-system interaction — Part 920: Guidance on tactile and haptic interactions

In the above mentioned standards, the standards listed at serial No. 3, 5 and 11 were also adopted by the Bureau of Indian Standards as parts of IS 16595. Part Numbers in IS 16595 have been retained the same part number as that of ISO. Other parts of this series may be adopted based on the recommendation of the technical committee in future.

This standard 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

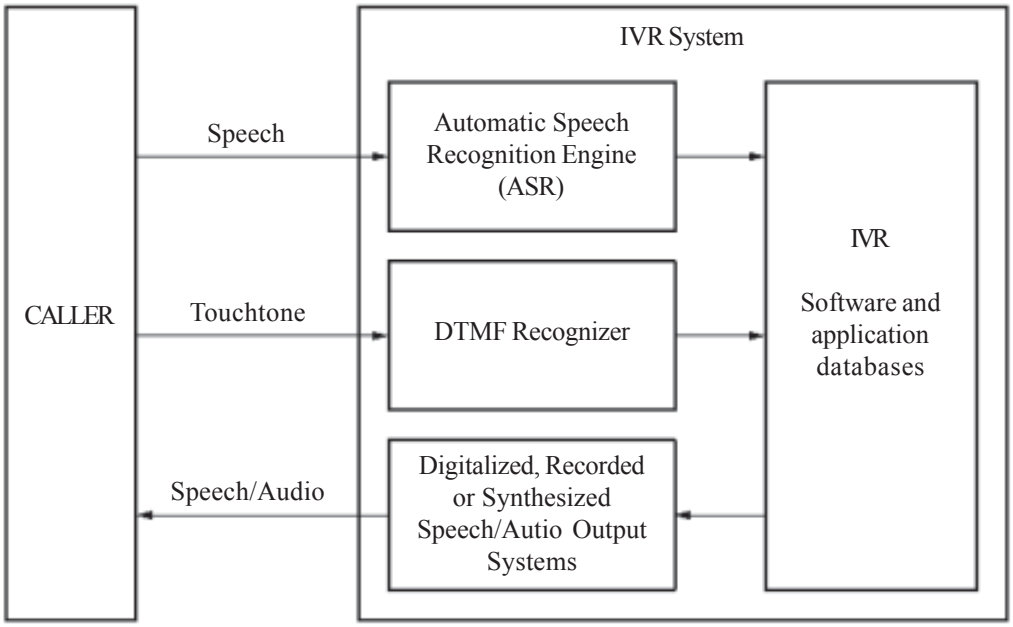

Fig. 1 IVR System Representation

the scope of this standard and are not addressed in this standard. 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 IS 16595 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 standard 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 standard 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.

This standard provides provisions for the design of IVR dialogues in speech-enabled and toucht one 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

standard. However, it 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 (for example, 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 (that is, applications that use statistical language models), detailed design provisions are not included for these applications in this standard 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.

The text of ISO Standard has been approved as suitable for publication as an Indian Standard without deviations. Certain conventions are however not identical to those used in Indian Standards. Attention is particularly drawn to the following:

a) Wherever the words 'International Standard' appear referring to this standard, they should be read as 'Indian Standard'.

b) Comma (,) has been used as a decimal marker while in Indian Standards, the current practice is to use a point (.) as the decimal marker.

The technical committee has reviewed the provisions of the following International Standards/other Publication referred in this adopted standard and has decided that are acceptable for use in conjunction with this standard:

International Standard	*Title*
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

Annex A and B in this standard are informative.

In reporting the result of a test or analysis made in accordance with this standard, is to be rounded off, it shall be done in accordance with IS 2 : 1960 'Rules for rounding off numerical-values (*revised*)'.

Indian Standard

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

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

3.15
discourse marker
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

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.

3.34
recognition error
general term for any error made by the ASR system during the recognition of speech

Note 1 to entry: Substitution errors, deletion errors, insertion errors and rejection errors are specific types of recognition errors.

3.35
recorded speech
stored human speech

Note 1 to entry: Contrast with *synthesized speech* (3.40), *digitized speech* (3.14) and *text-to-speech* (3.45).

Note 2 to entry: Sometimes called "canned speech".

3.36
rejection error
instance of a recognition error in which the ASR system does not recognize the spoken input although it is within the grammar

3.37
speaker dependent
speech recognition system in which distinct acoustic models are developed for each speaker using the system

3.38
state
<IVR> contextual properties of an IVR application that remain constant during a designated period of time

Note 1 to entry: An IVR application is often conceived as a set of distinct states and transitions to and from one state to another (cf. "finite state machines"). The current state defines the context of interaction and the valid caller input events. A state within an IVR is usually associated with a specific system output — see *message* (3.27) — the execution of internal system functions, a set of possible caller input events (i.e. legal DTMF input and/or an active vocabulary or grammar) and a set of transitions, triggered by events, which initiate movement to another specific state within the IVR.

[SOURCE: ISO 14915-3:2002, 3.6.11, modified.]

3.39
substitution error
instance of a recognition error in which a word (or sequence of words) in a speaker's utterance is incorrectly recognized as some other word or sequence of words in the recognizer's active vocabulary

3.40
synthesized speech
speech generated by a computer that is modelled computationally or constructed from the fragments of recorded human speech

Note 1 to entry: Contrast with *digitized speech* (3.14), *recorded speech* (3.35) and *text-to-speech* (3.45).

3.41
system initiative
<IVR> situation in which the IVR is leading the turn-taking with the caller

3.42
talk-ahead
capability in an IVR system of accepting spoken input before the system has requested it

Note 1 to entry: See *dial-ahead* (3.12).

Note 2 to entry: The spoken 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.

3.43
talk-through
capability in an IVR system of accepting speech input while a prompt (or an announcement) is being played

Note 1 to entry: In response to talk-through, playback of speech ceases and the system responds to the speech of the caller.

Note 2 to entry: See *dial-through* (3.13) and *barge-in* (3.3).

3.44
text telephone
TTY
telecommunication device which allows two-sided conversations to occur via text, as opposed to voice, communication

Note 1 to entry: In some environments this is referred to as TDD (telecommunication device for the deaf).

3.45
text-to-speech
TTS
process whereby a computer converts a representation of text into spoken, computer-generated speech

Note 1 to entry: The computer-generated speech is generally constructed from the fragments of recorded human speech or human speech that has been modelled computationally.

Note 2 to entry: Contrast with *synthesized speech* (3.40), *digitized speech* (3.14) and *recorded speech* (3.35).

3.46
time-out
interval of no user input that causes the system to change state; state change resulting from such an interval

EXAMPLE A time-out during numeric input may be interpreted as end-of-string and cause the system to change from a numeric input state to a command state or some appropriate action state.

3.47
tuning
process of analysing field data about the performance of a speech application and then using it to adjust ASR parameters in order to improve speech recognition performance

3.48
turn-taking
<IVR> process of the caller and the IVR alternating speaking and then listening

3.49
voice recognition
capability of a system to identify a specific person's voice

Note 1 to entry: This differs from speech recognition, which recognizes spoken words.

Note 2 to entry: Voice recognition is often used for speaker identification, authentication or verification in security applications.

4 Conformance

Conformance with this part of ISO 9241 is achieved by satisfying all the applicable requirements and by the provision of a systematic list of all the recommendations that have been satisfied. Any requirements that have been determined not to be applicable shall also be listed, together with a statement of the reasons why they are not applicable.

Users of this part of ISO 9241 shall evaluate the applicability of each requirement (a "shall" statement) and should evaluate the applicability of each recommendation (a "should" statement) to determine whether it is applicable in the particular context of use that has been established for the interactive

system that is being designed ("may" statements give permission). If a product is claimed to have met the applicable recommendations, the procedure used in establishing requirements for, developing and/or evaluating IVR user interfaces shall be specified. The level of specification of the procedure is a matter of negotiation between the involved parties.

5 Voice messaging systems

Voice messaging systems shall be in accordance with ISO/IEC 13714.

6 Information input

6.1 General

IVR systems are designed to accept speech input, touchtone input from the telephone keypad, or both. This clause contains provisions regarding caller input to IVR systems.

NOTE Users of text telephones (also called TTYs or TDDs) and voice relay services sometimes have special needs due to the constraints inherent in these assistive technologies and services. Provisions based on these constraints are also included in this section.

6.2 Informing text telephone users of acceptable input types

Applications should inform TTY users as to whether application input will be accepted via touchtones, text telephone, or both.

NOTE 1 Some text telephones do not have a touchtone input option and some IVR systems do not accept TTY input. Text telephone users need to be informed of the input requirements of the IVR system.

NOTE 2 A reasonable place in the dialogue for conveying this information is in the error messaging after a timeout or retry prompt for the application's first interactive state.

6.3 Non-duplication of information input

Within a given telephone call, unless information re-entry is required for reasons of privacy, security, or verification, the caller shall be required to enter any given piece of information only once.

NOTE This requirement also applies to all situations in which calls are transferred from one location to another as, for example, in call overflow situations. The caller is not expected to re-enter any information that has already been entered during the call.

EXAMPLE 1 If an application employs voice verification, a specific caller says his/her name or password twice to satisfy security requirements.

EXAMPLE 2 If a caller has entered his or her address in response to a prompt and the call is then transferred elsewhere for servicing, the address information is transferred with the call and the caller need not enter the information a second time.

6.4 Using system information to improve efficiency

If information is available in the system that would improve the dialogue with the caller, then it should be used to make the call more efficient.

EXAMPLE The credit card information for frequent fliers is stored so that they do not have to re-enter it for every reservation they make.

6.5 Changing information that has been entered

A caller should be provided with the opportunity to change data and information that have been entered during a call.

NOTE Sometimes there is a point in an application at which a caller can no longer make changes.

EXAMPLE After a caller has provided the information needed to complete a product order, the system repeats the order details and asks the caller for confirmation, providing an opportunity for the caller to make any desired changes before submitting the order. Once the caller has confirmed the order, however, the caller is unable to change it.

7 Speech input

7.1 Vocabulary choices for speech-enabled IVRs

When ASR systems are used, the performance of the ASR system shall be considered when selecting the vocabulary for the application and the proposed vocabulary should be tested.

7.2 Phonetically distinct vocabulary in speech-enabled IVR applications

For applications employing automated speech recognition, command options intended to be simultaneously available in response to an application prompt should be designed to be phonetically distinct, so as to facilitate speech recognition accuracy.

EXAMPLE In a voice mail application in English, the command option set “erase, repeat and next” is used, rather than “delete, repeat and next”, because “delete” and “repeat” are phonetically similar in English and might be confused by the ASR system.

7.3 State-specific grammar

Each dialogue state should have its own dedicated grammar.

NOTE 1 Even for common prompts, such as those requiring simple affirmative or negative responses, grammars are preferred because they facilitate handling of superfluous speech. A caller might respond to a yes/no question by saying “yes” or “no”, but also with speech like “oh…ok”, “yes, please”, “nope”, “absolutely not”, depending on the specific question asked and other factors.

EXAMPLE 1 A prompt requesting the caller to respond with the year their automobile was last serviced has a grammar that includes “two thousand nine”, “two thousand ten”, “twenty-ten” and “this year... so two thousand ten.” .

NOTE 2 Caller responses often contain parts of the preceding prompt.

EXAMPLE 2 In response to “Would you like to book that itinerary?”, callers might respond with, “Yes, I would.”; or “Book it.”

7.4 Synonyms in grammars

Grammars should include the prompt options and any likely word and phrase variants.

NOTE 1 Tuning is used to expand initial grammars based on field data.

NOTE 2 Creating grammars with all conceivable variants makes recognition more difficult and can lead to artificially low confidence scores.

EXAMPLE If a system presents the prompt “State your command: save, erase or review”, the grammar includes both “erase” and its synonym “delete”.

7.5 Use of dynamic grammars

Dynamic grammars should be used when speech recognition performance can be improved by constraining the grammar, typically when

a) the number of items in the grammar has become too large for acceptable speech recognition performance,

b) information about the caller is available to use for the grammar, or

c) information collected from previous caller input during a call is available to use for the grammar.

7.6 Enrolment

7.6.1 General

If enrolment is required in a speech-enabled IVR application, the process should be as simple as possible for the caller.

NOTE Sometimes there is a need to trade off ease of enrolment with the need for security. A more secure application often necessitates a longer, more arduous enrolment process.

7.6.2 Enrolment in the environment of use

If an application employs speech input and requires caller enrolment, it (or its documentation) should inform the caller that the enrolment should occur, if possible, in an environment similar to the one in which the application will be used.

7.6.3 Enrolment by repetition

If the system, as part of an enrolment procedure, instructs the caller to repeat certain phrases, each such phrase should be brief enough to be maintained easily in the caller's short-term memory. Specifically, if the caller is instructed to repeat sequences of digits, each repetition should not contain more than four digits.

7.7 End-pointing and turn-taking

End-pointing parameters should be set in a way that facilitates turn-taking.

EXAMPLE Many callers speak bank account or telephone numbers in chunks of several numbers with pauses in between, e.g. "<local code> – pause – <participant's number>". End-pointing parameters are set to reflect this method of input.

NOTE 1 In certain dialogue situations, caller input with natural speaking pauses can be expected. An end-pointing parameter set too short might cut off the caller before he or she has finished speaking.

NOTE 2 This subclause is related to 8.2, which addresses the use of explicit end-point delimiters for touchtone-based IVR systems.

NOTE 3 Pauses of less than 250 ms are unlikely to trigger turn-taking, whereas pauses exceeding 1 500 ms are very likely to trigger turn-taking. Thus it is reasonable to set end-pointing at about 1 500 ms, but special dialogue steps or user populations could justify the use of a larger value.

7.8 Handling time-outs for speech input

Because there is no speech equivalent to the pound ("#") key, a speech time-out value (separate from the touchtone time-out value) should be used to achieve a satisfactory system response time after caller input.

NOTE The separate speech time-out would need to account for natural pauses in speech.

7.9 Speech IVRs with touchtone

Speech input options shall have an equivalent means of touchtone input, unless the nature of the input does not lend itself to touchtone entry.

NOTE This requirement is particularly important for both usability and accessibility reasons.

EXAMPLE 1 The speech prompt, "Which would you like? A, B, or C" maps A to 1, B to 2 and C to 3 to accommodate touchtone input.

EXAMPLE 2 If an input error occurs at an open-ended dialogue prompt, then the IVR offers a hierarchically organized list of options, from which callers can choose using touchtone input.

8 Touchtone input

8.1 Indicating touchtone capability

If callers are explicitly asked to indicate they have a touchtone telephone by pressing a key, the prompt shall specify "1" as the appropriate key press, but any touchtone shall be recognized as a correct response to the request.

NOTE There may be equally valid ways of identifying touchtone capability other than requiring the caller to press a key. This requirement does not assume that this particular procedure represents the only or best way to determine whether or not a caller has a touchtone phone.

8.2 Key assignment for delimiters

If a key press delimiter is required for variable-length data entry, the pound ("#") key shall be used for this purpose.

NOTE In some countries other names are used for the pound ("#") key, such as "hash key".

8.3 Pound ("#") key in fixed length data entry

In applications employing touchtone input, use of the pound ("#") key as a delimiter is optional for fixed-length data entry; therefore, presses of the pound ("#") key during fixed-length data entry should be ignored.

8.4 Handling time-outs for variable length touchtone input

When receiving touchtone input from callers in applications employing variable length data entry, the pound ("#") key should be assumed automatically by the system after a developer-specified time-out period during which no caller input has been received. The length of the time-out period should be determined through user testing with a representative sample of callers, including those who could be expected to require relatively long intervals for the entry of data (e.g. the elderly, users with motor impairments, text telephone users).

NOTE This guideline implies that pressing "#" to delimit data entry is encouraged, but not required.

8.5 Key assignments for affirmative and negative responses

The key assignment for the affirmative response option shall always be 1. The key assignment for the negative response option shall always be 2.

EXAMPLE In an English-language application, "yes" and "no" responses are assigned to key 1 and key 2, respectively, on the touchtone keypad.

8.6 Key assignment for the human "help" function

When human help is available, both touchtone and speech-enabled applications shall assign the zero ("0") key for requests to reach a human.

NOTE 1 See also 12.2 and 12.7.

NOTE 2 This requirement does not preclude other keys or key combinations also being assigned to this function in addition to the use of zero ("0").

8.7 Minimization of caller key presses

In applications employing touchtone input, the number of key presses required of the caller should be minimized. Whenever feasible, single, rather than multiple, key presses should be expected of the caller.

9 Information output

9.1 General provisions for prompts and announcements

9.1.1 Information brevity

To avoid memory overload, information output should be brief.

NOTE Callers are able to mentally process only a small number of pieces of information (about four), such as dates, place names or numbers, in any one announcement or prompt, generally speaking. However, the exact number varies, depending on the complexity of the information and the degree of similarity among the items. Also, if the caller is not required to maintain all of the items in short-term memory, a number greater than four is sometimes justifiable.

9.1.2 Grouping information

When multiple pieces of information are presented in one announcement or prompt, pauses should be inserted between phrases to group information.

EXAMPLE Pauses are inserted between the phrases in the following announcement and each phrase presents one piece of information: "The flight number is 123. <break> It departs from gate 12. <break> Departure time is 3:45 pm."

9.1.3 Intonation and other aspects of prosody

To increase clarity, the prosody, including intonation patterns, of information items (such as numbers, digit strings and dates) should be similar to the prosody in human speech.

NOTE This is achievable by providing two or more representations of each digit or digit sequence, each with a different intonation, in the digitized, recorded or synthesized speech system used in the application.

EXAMPLE For English, when presenting the digit string "one two three", typically, the digits "one" and "two" have a rising/continuing intonation, whereas "three" has a falling/final intonation. If the sequence were "three two one", "three and two" would have a rising intonation, whereas "one" would have a falling/final intonation.

9.1.4 Speech rate

Speech rate for prompts and announcements should consider the user population, purpose of the IVR application and the dialect being used.

NOTE Speech rates often vary widely between IVR applications.

EXAMPLE A rate of 140 to 170 words per minute is used for customer service IVR applications in North America.

9.1.5 Simple linguistic forms

Simple linguistic constructions and word forms should be used whenever possible.

EXAMPLE 1 The prompt "Enter your account number now" is used instead of "An account number must be entered now" (passive voice).

NOTE In general, if there is a choice between a long and a short word meaning essentially the same thing, choose the short word. In most cases, the short word will be more common than the long word and callers will hear and process it more quickly. Use of well-constructed sentences and common vocabulary helps to ensure the largest number of callers will understand application messages and eases mental processing by the caller.

EXAMPLE 2 A prompt uses the word "use" instead of "utilize" and uses "help" instead of "assistance."

9.1.6 Judicious use of politeness language in input prompts

Politeness terms (e.g. "please, thank you") should be used with discretion within interactive voice response applications. Both overuse and under-use of such terms should be avoided.

NOTE 1 If polite words (e.g. "please") are used too often they can become tedious.

EXAMPLE If the system offers a list of menu options, the word "please" is not used before each option, but only at the beginning of the prompt (e.g. "Please select one of the following: for sales, press 1, for shipping press 2, for other, press 3.").

NOTE 2 The use of a politeness term at the beginning of a prompt (e.g. "Please enter your account number.") helps to prepare the caller to answer the request and may also positively dispose callers to the application and, indirectly, the company or organization whose application it is.

9.1.7 Disconnecting the call

If the IVR initiates disconnection of the call, then it should do so politely.

EXAMPLE "Thank you for calling the ABC Company. Goodbye."

9.1.8 Audio quality of prompts and announcements

The audio quality of prompts and announcements should be high. Ambient noise should be minimized during recording (i.e. background noise should be limited to the greatest extent possible).

NOTE 1 For callers with hearing impairments, separating relevant signals (e.g. prompts and announcements) from noise is sometimes very difficult.

NOTE 2 "High" audio quality is a subjective judgment of IVR application designers and stakeholders. These judgments are affected by distortion-free recording, clarity of speech and an understanding of the bandwidth and frequency limitations of phone lines.

9.1.9 Audio volume consistency of prompts and announcements

Prompts and announcements should be recorded and presented at a consistent audio volume throughout the application.

9.1.10 Caller perception of the IVR as a computer

The IVR application shall be constructed so that it is clear to callers that they are interacting with a computer, not a person.

EXAMPLE IVR system: "Welcome to Aladdin Company's automated service system."

NOTE 1 This requirement does not preclude the use of anthropomorphism in the persona, but anthropomorphism can make the compliance with this requirement more difficult.

NOTE 2 The use of non-speech audio also helps to convey that the caller is speaking to a computer.

9.1.11 Persona

The persona for an IVR application should be carefully chosen based on a systematic review and data collected from customers and should be used consistently throughout the IVR.

NOTE 1 All IVR applications actually have a persona (as defined in 3.31), whether the designers have constructed it or not. If the designers do not take an active role in constructing the persona, the application could be perceived by the target caller population(s) in a way not intended by the designers.

NOTE 2 Persona is not a replacement for good user interface design and is a secondary consideration after addressing the application's usability.

9.1.12 Repetition of prompts and announcements

a) Callers should be provided with a means of repeating prompts and announcements when they are very long or contain large lists of items from which callers will be asked to make a selection.

   NOTE Repeatability is not desirable for all prompts and announcements, and designers consider the various tradeoffs to determine the circumstances under which this functionality is appropriate.

b) In speech-enabled systems, "Repeat" shall be used as the speech command for requesting repetition of a prompt or announcement.

9.1.13 Repetition of portions of long prompts and announcements

If technically feasible, callers should be provided with a means of repeating portions of long prompts and announcements (e.g. a "short rewind" capability).

NOTE 1 Although the length that constitutes "long" for an announcement or prompt has not been specified in this document, application developers can determine through usability testing those instances in which a "short rewind" capability would be desirable because of the length or complexity of a prompt or announcement. It is important that such testing include representative users of the application.

EXAMPLE In an insurance claims application in which a caller is prompted to carry out a complex, multi-step instruction to complete a task and, subsequently, enter information into the system, a "short rewind" function is used, so the caller can verify a portion of the complex instructions they missed or forgot.

NOTE 2 The advisability, however, of having such long, multi-step instructions in an application is questionable.

9.1.14 Prompts and announcements for text telephone users

Prompts and announcements for text telephone users shall convey the same content as for non-text telephone users, but the wording needs to differ because the medium is different. Application developers should consult with text telephone users in the development of prompts and announcements for use with text telephones.

9.1.15 Language consistency

If the application allows the caller to specify the language used for prompts and the caller does so, all subsequent task-relevant system output shall be provided in the specified language.

NOTE If a caller has requested, for example, Spanish, and is put on hold with a radio station signal provided in the background, it would be appropriate to switch the radio to a Spanish-speaking station.

9.1.16 Language of text telephone (TTY) messages

The language of TTY messages shall be treated as a separate language.

NOTE This is because TTY messages involve abbreviations and certain conventions (e.g. in English "GA" to mean "go ahead") that other telephone communication systems do not. Consistent use of such abbreviations and conventions is important to text telephone users.

9.2 Construction of prompts and announcements

9.2.1 Ordering of items in option lists

Option lists should be ordered so that the highest priority options are presented first and any non-specific or miscellaneous options are presented at the end of the list. Factors such as frequency of use, granularity, revenue generation, logical order, importance to an upset caller and overlearning of the task model should be considered in determining the ordering of options.

NOTE If general options are presented first, callers tend to barge in and choose them, rather than wait for a specific option that better suits their needs. Placing frequently used options earlier in the list allows more callers to barge in quickly, improving the overall efficiency of the list.

EXAMPLE 1 "If this is an emergency, press 1 now. If you have a deadline that needs to be met, press 2. For regular service with delivery in three to seven days, press 3."

EXAMPLE 2 "For hotels in Europe, press 1. For hotels in Asia, press 2. For hotels in North America, press 3. For hotels anywhere else in the world, press 4."

9.2.2 Ordering of key assignments for option lists

Key assignments for option lists should be made in accordance with the ordering of list options and should be consecutive.

EXAMPLE If the list option order is "Account Balance," "Pay My Bill," "Close Account," and "Other," the key assignments are 1, 2, 3 and 4, respectively, for these options.

NOTE Universal keys/global commands are typically assigned high key values, e.g. "0" for human help, so as not to conflict with assigning items in the option list starting with key "1".

9.2.3 Ordering of goals and actions in prompts

The wording of input prompts should always first define the goal (result) and then the action necessary to achieve it.

EXAMPLE 1 "To place an order, press 1."

NOTE 1 This positions the action required of the caller as the most recently presented information to be remembered.

NOTE 2 Speech-enabled applications often combine the goal and action in the same word or phrase.

EXAMPLE 2 "Please say, either "new customer" or "existing customer."

9.2.4 Intonation and other aspects of prosody

To sound natural and to facilitate caller understanding, messages should be presented with proper prosody.

NOTE For recorded prompts, proper coaching of the voice talent is helpful for meeting this recommendation.

9.2.5 Intonation and other aspects of prosody for concatenated messages

For concatenated messages, it may be important to have three separate versions with different prosody, depending on the order of presentation of the individual messages: initial, medial and final.

NOTE See also 9.1.3.

9.2.6 Intonation and other aspects of prosody to facilitate turn-taking

System prompts should be presented with prosody that facilitates turn-taking, in order to make it easy for callers to know when it is their turn to speak. Most notably, prompts formed as questions should have an unambiguous question prosody.

NOTE For English, questions typically have a rising final intonation. For this reason, a short tag question like "Is that right?" is often used, as this phrase can easily be recorded with a clear rising intonation and prosody, indicating that the caller is expected to respond.

9.2.7 Preamble phrasing for speech option lists

Prompts should be worded such that a list of options does not initially sound like a yes/no question or a prompt for open-ended speech recognition.

EXAMPLE "And which of the following are you calling about: 'your bill', 'a payment', 'an order,' or 'technical support'?"

NOTE Prompts that are phrased like, "Are you calling about 'your bill', 'a payment', 'an order', or 'technical support'?" often cause a problem because callers barge in after "your bill" with a yes or no response. Or, a prompt such as: "Which service are you calling about: 'TV', 'Internet', or 'phone'" will elicit open-ended speech after "Which service are you calling about…"

9.2.8 Constraining speech input

An illustrative example should be provided where numerous spoken variants exist for the same information.

EXAMPLE "Please enter your credit card's expiration date. For example, 'December twenty twelve.'"

9.2.9 Use of discourse markers and landmarks in prompts

Prompts should begin with either a discourse marker or landmark, but consecutive prompts should not begin with the same word or phrase.

NOTE This helps to make dialogue with the IVR sound more natural and is effective in providing implicit feedback from the previous prompt and conveying a sense of progress toward accomplishing the caller's task.

9.2.10 Unambiguous prompts

Prompts should be stated so that they are unambiguous to the target caller population(s).

NOTE What is considered ambiguous is heavily dependent on the user population(s).

EXAMPLE Instead of a payment prompt saying "You have an outstanding balance of $ 50," where "outstanding" could be interpreted as a good situation or as an overdue balance, the prompt is worded as: "The balance due on your account is $ 50."

9.2.11 Incomplete set of options

List options should be constructed so that they contain only a single choice component. If a list option contains more than one choice component, then the caller should also be offered options with each of the individual choice components separately.

EXAMPLE "To order this item and charge it to your credit card, press 1" is an unclear option to the caller who wants to order the item, but NOT charge it to his or her credit card. At this point, the caller cannot know whether or not other options will be presented that will meet his or her needs. Instead, a better prompt asks the caller whether he or she wants to order, then an additional prompt requests information about the preferred method of payment.

9.2.12 Additional options

If all options are not presented, the caller should be notified and provided a means of identifying further options.

NOTE 1 In systems with a limited set of options, the most commonly used options are presented to the caller together with an option to get a listing of further options.

NOTE 2 In speech-based systems with a large set of options, the caller is presented with a limited set of example options.

9.2.13 Avoid explicit instructions on how to speak

In speech-based IVR systems, explicit instructions on how to speak should be avoided.

NOTE 1 Explicitly encouraging callers to speak naturally, fluently, etc. increases self-consciousness and as a consequence callers will speak even less naturally and less fluently.

NOTE 2 Speech recognition works best with clear, natural and fluent speech input by the user.

9.2.14 Grammar and terminology used in prompts

9.2.14.1 General

Terminology used in this section (such as *press* and *enter)* may be translated in order to take language-specific issues into account.

NOTE ETSI ES 202-076 provides equivalent terminology in numerous languages for many of the terms used here.

9.2.14.2 Vocabulary used in prompts

Prompts should contain only vocabulary that the expected caller population would be likely to understand.

NOTE A common mistake in designing IVR applications is the use of technical domain "jargon" which, while familiar to the designers, is not familiar to laypersons.

9.2.14.3 Consistency of terminology use

Terminology use should be consistent throughout an application and among applications that are integrated with each other within a system.

EXAMPLE If an IVR application, a messaging application and a paging application are linked together in one service, the terminology is used consistently throughout the three applications.

9.2.14.4 Location-specific variations in terminology

If there is a possibility that the caller will be transferred to or linked with different geographic locations when using a specific application, terminology use should be consistent among those locations, to the greatest extent possible.

NOTE Language use varies somewhat in different parts of a country and, provided that the vast majority of callers will be from a given geographic area, tailoring the application to accommodate local language variations is appropriate. However, such accommodations are problematical if a fair percentage of calls are likely to be transferred from one geographic location to another.

9.2.14.5 Use of the term "press" in prompts

The term "press" shall be used when the caller is asked to press a single key (e.g. to select an option from a presented list).

9.2.14.6 Use of the term "enter" in prompts

The term "enter" should be used in a prompt when the caller is asked to supply input requiring multiple key presses (e.g. to enter data such as an account number). The term "enter" should also be used in applications in which a single press of the pound ("#") key will accept the default.

EXAMPLE "To start service tomorrow, press 'pound'. Otherwise, please enter the desired service date."

NOTE 1 In effect, when a single press of the "#" key accepts the default, multiple key presses are implied (i.e. the default option plus the "#" key).

NOTE 2 Use of the term "dial" is sometimes used in prompts for telephone numbers.

9.2.14.7 Use of the term "record"

The term "record" should be used when prompting callers to leave a recorded message.

EXAMPLE "At the sound of the tone, please record your message."

9.2.14.8 Use of the terms "say", "state" and "select"

The terms "say", "state" and "select" should be used consistently in applications employing automated speech recognition (ASR) as the input mechanism.

EXAMPLE 1 "Please select 'reservations' or 'account information.'"

EXAMPLE 2 "Please state your name."

EXAMPLE 3 "To make a call, please say 'operator', 'person to person,' 'collect,' or 'third party.'"

NOTE There is some evidence to suggest that callers sometimes barge in after the first option, inappropriately, when "say" is used preceding an option list. Nevertheless, "say" is commonly used in many such prompts. Use of appropriate prosody in the prompt can counteract this problem.

9.2.14.9 Touchtone phone key names

The numeric keys on a touchtone phone shall be called by their numeric names and symbolic names in accordance with ITU-T E 161.

9.2.15 Irrelevant or unavailable options

If a selection in an option list becomes irrelevant or unavailable to the caller, based on the caller's previous choices or other circumstances pertinent to the application, that option should not be presented to the caller in any selectable list subsequently presented as part of the application.

EXAMPLE 1 If the system's automated bill payment function becomes unavailable to callers for a period of time, callers are not presented with the option of automated bill payment during this time period. In addition, they are informed early in the call that this function is not available: "Sorry, the bill payment option is temporarily unavailable. If you need to use this option, please call back at a later time."

EXAMPLE 2 If a caller to a travel arrangements IVR application has denoted a desire to travel by air, the application does not present rail options for travel.

EXAMPLE 3 An application's prompt option to speak with a human representative is not presented when no human representative is currently available.

9.2.16 Key assignments for unavailable/irrelevant options

Although unavailable/irrelevant options are not presented to callers (see 9.2.15), key assignments for these options should not be reassigned.

NOTE Given the tradeoffs, assigning options that are frequently omitted to the highest numbers in the menu is one way to make the omission less obvious to callers.

9.2.17 Consistency in key assignments

When feasible, the same key or speech command should be assigned to the same function throughout an application.

9.2.18 Consistency between speech and touchtone options

Touchtone option numbering should be consistent with the ordering of the speech options.

10 Navigation

10.1 Skip-forward function

If a system employing touchtone input provides a method for skipping forward to the end of an announcement, then the pound ("#") key should be used for this purpose.

EXAMPLE A caller chooses to hear only the title portion of each movie in a list of scheduled showings by pressing "#" as soon as the title is completed.

10.2 Dial-through and talk-through

Callers should be able to interrupt any prompt to give a response unless it is critical that they hear that information (e.g. certain error tones or messages). When a prompt is interrupted, any input that was received should be processed.

EXAMPLE 1 If the application prompts the caller for a telephone number and the caller interrupts the prompt and begins entering the number, the number is processed.

EXAMPLE 2 An application prompt that warns the caller that an action will result in an irreversible outcome (e.g. "Are you sure you wish to delete your unheard message?") is uninterruptible.

10.3 Dial-ahead and talk-ahead

Callers should be able to enter information that allows the system to process the information for subsequent steps without the callers being presented with any portion of the associated prompts.

NOTE Not all systems are capable of talk-ahead for technical or system performance reasons.

EXAMPLE To create a new greeting in a voice messaging system, callers respond to the first prompt by selecting "4" for greetings, then from the next prompt callers chooses "1" to access their standard greeting and then in the next prompt "2" to change their greeting. Dial-ahead is available and callers can press 4-1-2 in succession from the first prompt and proceed to record their greeting without being presented with the other two prompts.

10.4 Global commands

Other than those global commands used to reach a human representative and help, global commands should be used sparingly.

NOTE 1 For touchtone, there are a limited number of keys available for assignment to global commands. For speech IVRs, because the global commands are available at all times, each additional global command can degrade recognition performance for all prompts.

NOTE 2 For speech-enabled applications, some common global commands are "repeat," "start over," "go back," and "exit." It is important, however, to ensure that these commands do not acoustically conflict with local (i.e. regional-, country- or state-specific) commands.

10.5 Barge-in as a default

For speech IVRs, if the application is capable of barge-in, then it should be enabled by default.

NOTE See 10.3.

10.6 Restricting barge-in

Barge-in should be restricted

a) when it might cause the caller to miss critical information, and/or

b) in cases where callers inadvertently utter speech that is not relevant to the IVR at the opening of the call.

NOTE Because there is no expected caller input when announcements play, allowing barge-in leads to error conditions that are difficult to handle.

EXAMPLE 1 During the opening prompt for an application whose options have recently changed, barge-in is turned off so the caller can be informed of the changes.

EXAMPLE 2 While a caller is listening to terms and conditions for service during registration, barge-in is not available.

11 Help

11.1 Context-sensitive help

Help should be context sensitive, in that it should relate to the most recent entry made by the caller and any other available information that helps to infer the caller's task.

11.2 Caller-selectable help

If the system has help that is selectable at the caller's option, then

a) the help should be available at all points in the application,

b) the same command shall always be used to request help, and

c) the caller should be informed about how to request help.

11.3 Referencing help in initial prompt

Help should only be referenced in the initial prompt of an interaction if

a) the information requested is difficult for the caller to provide, and

b) a longer explanation will help significantly more callers provide the requested information compared to the number of callers who will be annoyed by the longer prompt.

NOTE Referencing help is not a good substitute for poorly worded prompts.

11.4 Help commands

Help may be offered as messages within the IVR or an option to connect with a human representative. If help is offered, then different commands (both touchtone and speech) should be assigned to the two methods for assistance.

EXAMPLE Commands such as “agent”, “operator”, “representative”, “customer service”, etc. and the zero key are the global commands for a live agent, whereas “help”, “more information”, etc. and the star (*) key are the global commands for IVR messages that contain help.

11.5 System-initiated help

After the caller encounters an error, the system may play a help prompt in place of a specific error prompt.

NOTE Often, the wording and content of error prompts (see Clause 14) and help prompts are similar.

EXAMPLE After two input errors by the caller, the system plays the help prompt for that state.

12 Access to human representatives

12.1 General

The caller should be able to access a human representative at any point in an application.

NOTE There are applications in which no human representative is accessible.

12.2 Equal access

If access to a human representative is provided, it shall be available to callers regardless of the input mechanism they use [e.g. pulse (rotary dial) phone, touchtone (i.e. dual-tone multi-frequency (DTMF) phone, text telephone (TTY) or automatic speech recognition (ASR)].

NOTE See also 8.6.

12.3 Prompting access

If access to a human representative is provided within the application, the caller should be given information that tells him/her how to access a human representative at contextually appropriate points in the dialogue.

NOTE 1 This is one of the most commonly user-requested features for IVRs. However, providing this information too early can increase the percentage of callers who opt out of the IVR.

NOTE 2 If an application is new to the caller or is very complex it is appropriate to give the information at an early point of the dialogue.

NOTE 3 It is appropriate to give information about access to a human representative if additional application-related information is frequently needed for choosing between options or if accessing a human representative is a major option.

EXAMPLE 1 After two consecutive input errors, the subsequent prompt adds the following option to the current list: "To speak to a customer service representative, say 'agent'."

EXAMPLE 2 After hearing the current status of an open order, the following prompt plays: "To discuss your order with a sales agent, press zero."

12.4 Initiating transfer

12.4.1 If the caller performs the action necessary to reach a human representative, the caller should be immediately transferred to a human representative or be put in a queue to reach a human representative and notified of this action (see, also, 13.1).

12.4.2 If the caller cannot be transferred or placed in a queue immediately, the system shall so inform the caller and give instructions as to what the caller can do next; the caller shall not be disconnected from the application or be presented with a "busy" in lieu of these instructions (see also 12.7).

EXAMPLE A caller using an IVR application at midnight to order articles of clothing presses "0" to access a human customer service representative. Human representatives are available, however, only during standard business hours. In response to the caller's command, the system replies: "We're sorry, but our customer service representatives are available only between 9:00 a.m. and 5:00 p.m. Monday to Friday. You can continue with your order by pressing '1' or call back during our regular business hours."

12.5 Delays in access

Unless continuous feedback is being provided to the caller, the caller should receive, at a minimum, a voice message after each 20 s period that elapses subsequent to the caller being placed in the queue. These voice messages should be varied so that the caller does not hear exactly the same message at the end of each waiting period. In text telephone user interfaces, text messages should be presented at frequent intervals and should also be varied from one presentation to the next.

EXAMPLE In a text telephone user interface, the initial delay message might be "PLS HD" (please hold). After 20 s, the character string "*****" could be sent, followed by "/////" after 40 s, etc.

NOTE 1 Continuous feedback (e.g. music on hold) is a reasonable alternative to intermittent voice messages in this situation. There is no consensus as to which is better.

NOTE 2 Messages that further inform the caller regarding the number of people in the queue, the estimated length of the delay or alternatives to holding are desirable.

NOTE 3 Voice messages are provided intermittently to the waiting caller so that he or she will be certain that some action is in process with respect to the request to speak with a human representative.

NOTE 4 See also 13.9, 13.10 and 13.11 for information about handling other types of delay situations.

12.6 Assistance from human representatives

If available, callers should be transferred automatically to a human representative at any point at which a pre-determined threshold is exceeded in terms of speech-recognition errors or other measures of caller frustration.

EXAMPLE When the caller exceeds a predetermined number of errors at any one point in the IVR or cumulatively during the course of the call, they are transferred to an agent to complete their request.

12.7 Human help not available

If the IVR application ordinarily offers access to human help, but current circumstances prevent access to this form of assistance, then the system should inform the caller that human help is unavailable and indicate the options available to the caller at that point.

13 Feedback

13.1 General

Feedback is an important means of keeping the conversational dialogue flowing in an IVR application. It serves, among other things, to inform the caller of the consequences of his or her actions and to communicate information about what the caller can do next. Feedback is not always delivered explicitly in the form of a specific feedback message from the IVR system, but is sometimes conveyed in an implicit manner by the subsequent actions of the IVR system. This clause provides provisions for the design of feedback in IVR applications.

NOTE Confirmations are one form of feedback to the caller. See 13.7.

13.2 Feedback to caller input

Feedback, either implicit or explicit, should be provided to the caller after every input, so the caller knows that the input has been received by the system.

NOTE See also 13.7.3.

EXAMPLE 1 The next list of options is presented immediately if the caller has responded to the previous one.

EXAMPLE 2 After entering an order number, the system repeats the order number for the caller to confirm.

EXAMPLE 3 The message, “Please wait while we access your account information”, is presented in response to caller entry of an account number.

EXAMPLE 4 After a caller selects “billing” from a list of options, the next prompt begins, “Okay, billing…”

13.3 Feedback following selection of unavailable options

If a caller selects an unavailable option, the IVR system shall provide feedback indicating that the option is not available.

NOTE Callers using dial-through, talk-through or barge-in are apt to select unavailable options more frequently because they do not listen to all list options before responding. Thus, they will need feedback to inform them that an option has become unavailable.

13.4 System response time

The IVR should quickly respond with its next message after caller input is completed.

NOTE If the application delays its response, then callers tend to think that the IVR did not hear them or that more information is required, which leads to increased input errors.

13.5 Appropriate context for feedback messages

Feedback messages should be stated in context-relevant terms.

EXAMPLE 1 Appropriate terms such as “euros” and, perhaps, “cents” are used for currency.

EXAMPLE 2 In English “A.M.” and “P.M.” are used for 12 h time, and noon and midnight are used to differentiate 12:00 p.m. and 12:00 a.m., respectively.

13.6 Landmarks

Landmarks may be used within the call flow to continue to orient the caller.

EXAMPLE The announcement "main menu" is played at the beginning of the application's central list of options, both for callers who first enter the application and for those who are returning from other parts of the application.

NOTE 1 The repetitive use of landmarks sometimes becomes tedious to the caller and, particularly in speech-enabled applications, is not conducive to a conversational experience.

NOTE 2 The content of a prompt often provides enough context to make a landmark superfluous.

NOTE 3 Landmarks are especially useful for applications that have significant recursive use.

13.7 Confirmation dialogues

13.7.1 General

Confirmation of input is sometimes desirable, if not necessary, in many IVR applications. In speech-enabled IVR systems, input confirmation is of special importance, as automatic speech recognition is error-prone. Different confirmation strategies are available, depending on the dialogue context. The following subclauses contain provisions specific to input confirmation.

13.7.2 No excessive confirmation

For speech applications, confirmation of caller input should not become excessive.

NOTE Although it can be used to improve confidence that the IVR is correctly recognizing the caller's speech, repeated confirmations slow down the call, annoy the customer and give him or her the impression that the recognizer is not reliable. Implicit confirmation can provide a solution to this problem.

EXAMPLE When collecting several items of information about a payment (e.g. amount, date, credit card number, expiration date, security code), only the credit card number is confirmed; the remainder of the information is confirmed as a batch near the end of the payment process.

13.7.3 Confirmation of simple words and phrases

Unless the consequences are irreversibly negative, simple utterances should not be confirmed explicitly.

NOTE Speech input such as "yes", "no", "help" and "repeat" are simple enough and confirming them conveys a very negative view of the speech recognition.

13.7.4 Explicit input confirmation to avoid detrimental system actions

The IVR system shall ask callers to explicitly confirm data entry whenever the consequences of not doing so could be unrecoverable and/or irreversibly negative. Explicit confirmation should be considered whenever the consequences of the caller's data entry could be detrimental to the caller, even if temporary and recoverable.

EXAMPLE 1 A caller confirms the dollar amount entered during a bill payment transaction to ensure that the correct amount was credited to the correct account.

EXAMPLE 2 The following dialogue exemplifies a confirmation situation: "You have requested to cancel your booking on Flight 303. If this is correct, press 1. Otherwise, press 2."

13.7.5 Confirmation of speech input based on confidence scores

13.7.5.1 If recognition of a speech input has a high confidence score, explicit confirmation is not required; implicit confirmation can be used, however.

13.7.5.2 If recognition of a speech input has a medium confidence score, an explicit confirmation dialogue should be used.

13.7.5.3 If recognition of a speech input has a low confidence score, confirmation is not appropriate and the input should be rejected and the caller reprompted.

NOTE Classifying speech recognition confidence as high, medium and low is an effective strategy for minimizing the overhead of confirming speech input. Using this strategy means that only input with confidence scores of medium is confirmed explicitly.

13.7.6 Combining confirmations

When several related pieces of information are collected consecutively, one batch confirmation may be used.

EXAMPLE In a payment task, the system confirms multiple pieces of information with: "Okay, so you're making a $ 50 payment today with your credit card ending in 1234, correct?"

13.7.7 Preambles for confirmation prompts

Context-specific confirmation messages and apology preambles should be used.

NOTE Some systems omit preamble messages altogether.

EXAMPLE Medium confidence speech is confirmed with, "Was that…?"; confirmation messages start with, "that was…, correct?"; and a prompt in response to a second recognition error begins, "My mistake again….".

13.7.8 Wording for confirmation prompts and announcements

Confirmation prompts should focus on the correctness of the information being confirmed, rather than the exact spoken or keyed utterance. Furthermore, whenever possible and non-ambiguous, the meaningful part of a confirmation prompt should be worded similarly to the caller's input.

EXAMPLE If the caller's utterance is: "March 21st", it is better to confirm this as "March 21st, is that correct?", rather than "I think you said 21st of March, is that correct?" The latter example does not repeat the caller's input in the same form.

13.7.9 Confirmation of digit or number strings

Whenever a number string is of a data type that has a conventional spoken format, confirmation prompts and announcements should be worded accordingly.

NOTE Because number string input is prone to automated-speech-recognition errors and often constitutes critical information within the IVR application, confirmation is desirable, especially when ASR is used.

EXAMPLE Swedish social security numbers, which consist of 10 digits, are customarily spoken in groups of two digits read as numbers, e.g. "seventy-five, twenty-one, thirty, sixteen, eighty". Confirming them in a different format, for example, digit by digit – "seven, five two, one, …", could confuse callers.

13.8 Intelligent use of confirmation rejections

If the caller has eliminated a choice during confirmation, that choice should not be offered again.

NOTE 1 This avoids situations in which the IVR repetitively tries to confirm an option the caller has already explicitly said they did not want.

NOTE 2 Option lists that delete choices a caller has previously eliminated by virtue of his or her prior selections are commonly referred to as "skip lists."

13.9 Additional feedback in delay situations

Feedback should be provided to the caller during any waiting period with a long duration so that the caller knows that the input is being acted upon. Text telephone users should also be served in this way via periodic text messages (see 12.5).

NOTE Feedback can be provided by music and/or announcements.

EXAMPLE After a caller provides an airline reservation number, there is a delay during which music is played while the system looks up the reservation.

13.10 Managing extended hold situations

Requests to continue holding should be repeated with some variation in the message if the caller's wait is long. Text telephone users should also be served in this way via periodic text messages. See also 12.5.

EXAMPLE After a two-second wait: "We are processing your order." Seven to 10 s thereafter: "Please hold. We are still processing your order." Additional variations are used if the hold situation continues for some additional period of time.

13.11 Suggesting that the caller call back

If it is not possible to process the caller's request within a time period acceptable to the caller, the system should inform the caller of the problem and the caller's options or alternatives.

EXAMPLE In the event that a system used for order processing has become inoperative, a message such as the following could be conveyed to the caller: "We are unable to access our computer ordering system at this time. You may FAX your order to us at [phone number], remain on the line to be connected with a customer service representative or try our automated ordering system again later."

14 Errors

14.1 General

There are many types of errors that occur in IVR systems. Input entry errors are relatively infrequent in touchtone-based IVR systems, but occur quite often in speech-enabled IVR systems, due to recognition errors of speech input by the ASR system. Many other types of errors also occur that are related to the cognitive aspects of the tasks callers perform using an IVR system and the structure and navigational aspects of the IVR application's design, among other things. This clause contains provisions for design relevant to the prevention and management of errors in IVR applications.

NOTE Confirmation is sometimes considered as a method of handling errors. See 13.7.

14.2 Unavailability of service

If callers select an option after which no subsequent options are valid, they should be informed and provided with further information concerning what to do next.

EXAMPLE If the caller selects a service option that is not available at that particular time (e.g. because of system malfunction), the system can so inform the caller and request that he or she leave a message or call back at a later time.

14.3 Input prompt repetitions

At any given point in the dialogue, an identical input prompt should be provided to the caller no more than twice in succession.

EXAMPLE If a time-out occurs after a caller is asked "Please enter your credit card number now", the identical prompt can be used again to solicit the caller's credit card number. If a second time-out occurs after the prompt repetition, the next prompt might be "Your credit card number is on the top right-hand corner of your bill. Please enter it now."

14.4 Error message content

If the caller enters invalid input in response to a prompt, the system should present a brief feedback response to the caller. If the caller's subsequent response is also invalid, more explicit feedback should be provided that helps the caller understand how to enter valid information.

EXAMPLE If the caller enters nine instead of the required 10 digits for an account number, the message "Our system recognized only nine digits in your account number. Please re-enter the number." would be a reasonable response to the error. If the caller re-enters the account number and makes yet another error, an appropriate response might be "Please enter the 10 digit account number printed on the top-right corner of your bill."

14.5 Two consecutive input entry errors in touchtone-based IVRs

If the caller has heard the same substantive input prompt twice and has failed to respond or provided an invalid response both times, the caller should be

a) transferred to a human representative, if available,

b) presented with a third prompt containing additional relevant information, or

c) transferred to an exit routine within the application (see 14.12).

14.6 Two or more consecutive input entry errors in speech-enabled IVRs

Because speech-enabled IVR applications are more prone to errors due to the use of ASR systems, callers who have failed to respond or provided an invalid response both times in response to the same substantive prompt should not be transferred to a human representative at that point. Additional prompting should be employed, depending on the speech recognition data, to attempt to determine and manage the problem before proceeding to transfer the call to a human representative or proceeding to an exit routine.

14.7 Handling of numerous errors

Numerous errors in a state should trigger alternative handling.

NOTE Repeating substantially the same prompt more than a few times often does not repair the dialogue with the caller. If errors become numerous, then other techniques for breaking the impasse are warranted — such as routing to an agent or bypassing the current prompt.

14.8 Avoiding pejorative wording in error messages

Words such as "error", "wrong", "invalid" and "illegal" should be avoided in error messages.

EXAMPLE In the case of an error in the entry of numerical data, the caller might receive the message, "We do not a have a record for this number. Please check the number and enter it again."

14.9 Critical messages

If it is critical that the caller hear specific information in an error message or informational message, that part of the message should be uninterruptible (barge-in should be switched off), but also as brief as possible.

EXAMPLE If the caller is attempting to page someone and the page did not go through, a short, uninterruptible feedback message to the caller is appropriate.

NOTE Because these messages are uninterruptible, it is important, from the caller's perspective, that their use be limited to the most critical situations.

14.10 Error recovery

Except as specified in 14.5 and 14.7, callers should be provided with the opportunity to recover from the most recent error.

14.11 Disconnection messages

Under normal operating conditions, the caller shall not be disconnected from the application until an explicit exit message has been presented.

14.12 Exiting due to exceeding error limit

The exit routine in the case of excessive errors should consist of an explanation of the problem that caused the error, suggestions for a possible resolution of the problem (when appropriate) and a polite termination unless the call can be transferred to a human representative.

EXAMPLE "Sorry, we cannot find information for that account number and there is no one available to assist you at this time. For further assistance, please call xxxxxx-xxxx during standard business hours. Thank you for calling. Goodbye."

14.13 Referencing help

Context-sensitive help or information on how to get help should be provided in error prompts.

14.14 Error prompt wording

Common sources of errors should be considered when writing error reprompts.

14.15 Cueing the caller about desired speech input

In speech-enabled systems, error messages for unrecognized speech input should be clear about the exact words or phrases the recognizer is looking for and should not simply repeat the initial prompt.

NOTE This recommendation is primarily for prompts with limited grammars. For prompts that intend to elicit natural language, using different examples might achieve this same goal.

14.16 Conversational repair

For speech dialogues, opportunities for conversational repair should be utilized by the IVR system, if feasible.

EXAMPLE 1 When rejecting a confirmation prompt (e.g. "I think you said Monaco, is that correct?") many callers will volunteer the correct information as well (e.g. "No, Bamako"). The system can achieve conversational repair by using the information given by the caller and can expedite the turn-taking, in this case by recognizing "Bamako" as the caller's intended input and proceeding to confirmation of that input ("Sorry. Bamako, right?"), rather than simply reprompting the caller to input the city name again (e.g. "Please state your destination.").

EXAMPLE 2 A caller ordering clothing using a speech-enabled IVR system has input her desire to order two shirts. The IVR system misrecognizes the word "two" and replies "You are ordering three shirts, right?" The caller replies, "No, only two." The IVR recognizes the "two" correctly this time and responds "I'm sorry, you're ordering two shirts. What is the size of the first shirt?", rather than responding with "I'm sorry, how many shirts did you want to order?"

14.17 Handling too much speech input

For open-ended speech prompts, a special uninterruptible prompt should be used when the caller speaks an utterance that exceeds the allowed maximum duration.

NOTE 1 The allowed maximum duration of speech would typically be between 5 and 15 s, depending on the context of an open-ended speech prompt.

NOTE 2 The justification for having a non-interruptible error prompt is that if the caller continues talking, the error prompt will be interrupted and the caller will not receive the information required to correct the error.

14.18 Mentioning touchtone options for speech IVRs

The touchtone options should be included in error prompts, especially after a recognition error.

EXAMPLE After a second recognition error, the application prompts: "I'm sorry, I still didn't understand. You can say 'billing' or press one, 'payments' or press two, 'sales' or press three, or 'technical support' or press four. For more information, say 'help' or press star."

Annex A
(informative)

Overview of the ISO 9241 series

The annex presents an overview of the structure of ISO 9241. For an up-to-date overview of its structure, subject areas and the current status of both published and projected parts, please refer to:

ISO 9241 series

The structure reflects the numbering of the original ISO 9241 standard; for example, displays were originally Part 3 and are now the 300 series. In each section, the "hundred" is an introduction to the section; for example, Part 100 gives an introduction to the software-ergonomics parts. See Table A.1.

Table A.1 — Structure of ISO 9241, *Ergonomics of human–system interaction*

Part	Title
1	Introduction
2	Job design
11	Hardware and software usability
20	Accessibility and human–system interaction
21–99	Reserved numbers
100	Software ergonomics
200	Human–system interaction processes
300	Displays and display-related hardware
400	Physical input devices — Ergonomics principles
500	Workplace ergonomics
600	Environment ergonomics
700	Control rooms
900	Tactile and haptic interactions

Annex B
(informative)

The role of speech recognition errors in IVR design

B.1 Recognition errors are inherent in speech-based IVRs

IVRs which accept speech input rely on an automatic speech recognition system (ASR) to interpret the caller's input. As with any other recognition technology, ASR is prone to errors. Despite recent and ongoing improvements in ASR performance, it will never be possible to totally eliminate recognition errors in speech-based IVR systems. The reason for this is to be found in the statistical process that underlies today's ASR systems.

When finite-state grammar applications are used, for every caller utterance a classification process assigns probabilities to all possible recognition results which have been predefined in the grammar of the specific IVR. As a result, the ASR engine returns a list of the *n*-best candidates, each together with a numeric score (confidence score) representing the hypothesized probability of correctness.

In order to make an adequate decision on how to continue with the dialogue, the system uses predefined confidence thresholds in order to interpret the scored *n*-best list:

— If all confidence scores lie below a given threshold, the system will react with a rejection, i.e. with no other obvious action than a message such as "Sorry, I could not understand you. Please repeat."

— A high-confidence score for one candidate and significantly lower scores for the other candidates will result in a triggering of the respective system action.

— Some IVR systems process medium confidence levels by presenting disambiguation dialogues, in order to clarify the caller's input.

B.2 Types of recognition errors and their consequences

B.2.1 General

Table B.1 provides an overview of possible combinations of caller utterances and ASR results. Recognition errors are highlighted by bold fonts.

Table B.1 — Overview of possible combinations of caller utterances and ASR results

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="4">ASR result</td></tr>
<tr><td>No input</td><td>“a”</td><td>“b”</td><td>No match</td></tr>
<tr><td rowspan="4">Caller input</td><td>No utterance</td><td>Correct timeout</td><td colspan="3">Insertion error</td></tr>
<tr><td>Valid utterance “A”</td><td rowspan="3">Deletion error</td><td>Correct recognition</td><td>Substitution error</td><td rowspan="2">Rejection error</td></tr>
<tr><td>Valid utterance “B”</td><td>Substitution error</td><td>Correct recognition</td></tr>
<tr><td>Invalid utterance</td><td colspan="2">False recognition</td><td>Correct rejection</td></tr>
</table>

B.2.2 Correct timeout

If the system does not recognize any caller input for a predefined period of time, a timeout or "no input" event is recognized which triggers a defined system behaviour (e.g. a message such as "Sorry, I did not hear you …").

B.2.3 Insertion error

If the caller does not make any speech input, but the IVR recognizes some type of speech input, then an insertion error occurs. This type of error often occurs when background noise in the caller's surroundings or non-speech sounds such as coughing, clearing of the throat, etc. are misprocessed. Insertion errors often lead to confusion, as the caller cannot understand the reasons for the system reaction. Insertion errors can have severe consequences for dialogue success and caller acceptance, especially for inexperienced callers and when the insertion error occurs at the very beginning of the dialogue.

B.2.4 Deletion error

Deletion errors occur when the caller makes an utterance, but the IVR does not register input, and include instances when barge-in is not triggered. Deletion errors also include instances where only a part of the caller utterance is incorrectly omitted in the speech recognizer's output. Reasons for deletion errors can be a mismatch of ASR sensitivity and the intensity of (certain parts of) the caller input or end-pointing errors that lead to incorrect recognition of the speech onset or offset.

B.2.5 Substitution error

A substitution error, sometimes called "misrecognition", occurs when the caller speaks an in-grammar option, but a different in-grammar option is recognized. The consequences of these errors are especially negative and more so if the IVR user does not realize that there has been a problem with the recognition. Substitutions can lead to unintended transactions (e.g. wrong bank account for a bank transfer) and misinformation.

B.2.6 Rejection error

A rejection error occurs when the caller provides in-grammar speech, but rejection usually occurs due to low confidence scores. This type of error is also known as a "no match" error. In comparison to substitution errors, rejection errors are directly obvious to the caller. In case of a rejection, the system is aware of a problem and communicates the problem to the caller which can be a good starting point for a cooperative problem resolution. Callers may become annoyed by repeated rejection errors. These errors also hinder efficient task completion and can lead to a negative opinion about the system's ASR performance.

B.2.7 False recognition

A false recognition occurs when an out-of-grammar utterance is given by the caller, but the system recognized an in-grammar option. Effects of false recognitions lie somewhere between those of insertions and substitutions. From the caller's perspective, the system's reaction in such cases seems to be arbitrary.

B.2.8 Correct rejection

If the caller gives an out-of-grammar utterance and the system rejects it, this is a correct rejection. While this is not an error on the part of the IVR system, dialogue strategies that encourage in-grammar utterances are needed to minimize this type of outcome.

B.2.9 Frequency of error types

In terms of frequency and effects on the usability of a speech-based IVR, substitutions and false rejections are the most important error types, because of the mismatch between the caller's utterance and the system's recognized option. The relative frequencies of both error types can be easily influenced

by manipulating confidence thresholds. Higher confidence thresholds will help to reduce substitution errors, but will result in relatively higher rates of rejection errors. However, lower confidence thresholds will decrease the number of rejections, but increase the number of substitutions. These mechanisms are often used in order to adapt the ASR performance to application-specific requirements of an IVR.

B.3 Voice user interface design for managing recognition errors

Considering possible recognition errors is a major issue for the ergonomic design of speech-enabled IVR user interfaces. Appropriate speech user interface design can help to

a) avoid recognition errors or minimize their frequency by

   — careful selection of valid caller utterances (see the guidelines on vocabulary and grammar design given in Clause 7), and

   — prompting that supports and guides the caller in producing intact, in-grammar speech input (see the guidelines on information output and help given in Clauses 9 and 11), and

b) reduce the negative consequences of recognition errors by

   — supporting the caller in detecting recognition errors (see the guideline on feedback to caller input in 13.2),

   — avoiding recognition errors that result in detrimental system actions (see the guidelines on confirmation dialogues in 13.7),

   — supporting the caller in recovering from recognition errors (see the guidelines of Clause 14),

   — establishing caller acceptance for the imperfection of speech by adequate wording in error and help prompts and feedback messages (see the guidelines given in Clauses 11, 13 and 14), and

   — providing access to a human (see Clause 12).

Bibliography

[1] ANSI/HFES 200.4, *Human Factors Engineering of Software User Interfaces — Part 4: Interactive Voice Response (IVR) and Telephony*

[2] ISO 9241-920:2009, *Ergonomics of human-system interaction — Part 920: Guidance on tactile and haptic interactions*

[3] Baber C., & Noyes J.M.eds. *Interactive speech technology: Human factors issues in the application of speech input/output to computers*. Taylor & Francis, Bristol, PA, 1993

[4] Balentine B., Morgan D.P. (2001). (2ND Ed.). *How to build a speech recognition application*. San Ramon, CA: Enterprise Integration Group, Inc.

[5] Bernsen N.O., Dybkjaer H., Dybkjaer L. *Designing interactive speech systems: From first ideas to user testing*. Springer Verlag, New York, 1998

[6] Cohen M.H., Giangola J.P., Balogh J. *Voice user interface design*. Addison Wesley, Boston, 2004

[7] Commarford P.M., Lewis J.R., Smither J.A., Gentzler D. A comparison of broad versus deep auditory menu structures. *Hum. Factors*. 2008, **50** pp. 70–89

[8] Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav. Brain Sci.* 2001, **24** pp. 87–185

[9] Deffner G.P., Melder K. (1990). User acceptance and preference for advanced voice services features and dialogue styles. Proceedings of the Human Factors Society 34th Annual Meeting (pp. 194-197). Santa Monica, CA: Human Factors Society

[10] ETSI ES 202-076 v.2.1.1:2009-08, *Generic spoken command vocabulary for ICT devices and services*

[11] Fay D. (1993). Interfaces to automated telephone services: Do users prefer touchtone or automatic speech recognition? In Proceedings of the 14th International Symposium on Human Factors in Telecommunications (pp. 339-349). Darmstadt, Germany: R.v. Decker's Verlag

[12] Frankish C., & Noyes J. Sources of human error in data entry tasks using speech input. *Hum. Factors*. 1990, **32** (6) pp. 697–716

[13] Gardner-Bonneau D.J., & Blanchard H.E.eds. *Human factors and voice interactive systems*. Springer, Boston, 2008

[14] Harris R.A. *Voice interaction design*. Morgan Kaufmann, Amsterdam, 2005

[15] Kortum P.ed. *HCI beyond the GUI: Design for haptic, speech, olfactory, and other non-traditional interfaces*. Morgan Kaufmann, Amsterdam, 2008

[16] Halstead-Nussloch R. (1989). The design of phone-based interfaces for consumers. Proceedings of CHI-89: Human Factors in Computing Systems (pp. 347-352). New York: ACM

[17] Han S.H., Williges B.H., Williges R.C. (1991). Quantitative guidelines for telephone information systems. Proceedings of the Human Factors Society 35th Annual Meeting (pp. 225-229). Santa Monica, CA: Human Factors Society

[18] Jones D., Hapeshi K., Frankish C. Design guidelines for speech recognition interfaces. *Appl. Ergon.* 1989, **20** (1) pp. 47–52

[19] Knott B.A., Bushey R., Martin J.M. (2004). Natural language prompts for an automated call router: Examples increase the clarity of user responses. Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting (pp. 736-739). Santa Monica, CA: Human Factors and Ergonomics Society

[20] McCauley M.E.Human factors in voice technology. In: *Human factors review*, (Muckler F.A.ed.). Human Factors Society, Santa Monica, CA, 1984, pp. 131–66.

[21] Marics M.A., & Englebeck G.Designing voice menu applications for telephones. In: *Handbook of human-computer interaction*, (Helander M.G., Landauer T.K., Prabhu P.eds.). Elsevier, New York, Second Edition, 1998, pp. 1085–102.

[22] Martin M.M., Williges B.H., Williges R.C. (1990). Improving the design of telephone-based systems. Proceedings of the Human Factors Society 34th Annual Meeting (pp. 198-202). Santa Monica, CA: Human Factors Society

[23] Marx M., Schmandt C. (1996). MailCall: Message presentation and navigation in a non-visual environment. Proceedings of CHI-96: Human Factors in Computing Systems (pp. 165-172). New York: ACM

[24] Miller M.A., Elias J.W. (1991). Using menus to access computers via phone-based interfaces. Proceedings of the Human Factors Society 35th Annual Meeting (pp. 235-239). Santa Monica, CA: Human Factors Society

[25] Reeves B., & Nass C. *The media equation*. Cambridge University Press, New York, 1996

[26] Resnick P., Virzi R.A. (1992). Skip and scan: Cleaning up telephone interfaces. In Proceedings of CHI-92: Human Factors in Computing Systems (pp. 419-426). New York: ACM

[27] Rodman R.D. *Computer speech technology*. Artech House, Norwood, MA, 1999

[28] Rosenfeld R., Olsen D., Rudnicky A. Universal speech interfaces. *Interaction*. 2001, **8** pp. 34–44

[29] Schmandt C. *Voice communication with computers: Conversational systems*. Van Nostrand Reinhold, New York, 1994

[30] Schumacher R.M. (1992). Phone-based interfaces: Research and guidelines. Proceedings of the Human Factors Society 36th Annual Meeting (pp. 1051-1055). Santa Monica, CA: Human Factors Society

[31] Schumacher R.M. Jr., Hardzinski M.L., Schwartz A.L. Increasing the usability of interactive voice response systems: Research and guidelines for phone-based interfaces. *Hum. Factors*. 1995, **37** (2) pp. 251–264

[32] Stuart R., Desurvire H., Dews S. (1991). The truncation of prompts in phone-based interfaces: Using TOTT in evaluations. Proceedings of the Human Factors Society 35th Annual Meeting (pp. 230-234). Santa Monica, CA: Human Factors Society

[33] Syrdal A., Bennett R., Greenspan S.eds. *Applied speech technology*. CRC Press, Boca Raton, FL, 1995

[34] Virzi R.A., Huitema J.S. (1997). Telephone-based menus: Evidence that broader is better than deeper. Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting (pp. 315-319). Santa Monica, CA: Human Factors and Ergonomics Society

[35] Virzi R.A. (1991). A preference evaluation of three dialing plans for a residential, phone-based information service. Proceedings of the Human Factors Society 35th Annual Meeting (pp. 240-243). Santa Monica, CA: Human Factors Society

[36] Virzi R.A., Resnick P., Ottens D. (1992). Skip and scan telephone menus: User performance as a function of experience. Proceedings of the Human Factors Society 36th Annual Meeting (pp. 211-215). Santa Monica, CA: Human Factors Society

Bureau of Indian Standards

BIS is a statutory institution established under the *Bureau of Indian Standards Act*, 1986 to promote harmonious development of the activities of standardization, marking and quality certification of goods and attending to connected matters in the country.

Copyright

Review of Indian Standards

Amendments are issued to standards as the need arises on the basis of comments. Standards are also reviewed periodically; a standard along with amendments is reaffirmed when such review indicates that no changes are needed; if the review indicates that changes are needed, it is taken up for revision. Users of Indian Standards should ascertain that they are in possession of the latest amendments or edition by referring to the latest issue of 'BIS Catalogue' and 'Standards: Monthly Additions'.

This Indian Standard has been developed from Doc No.: PGD 15 (1513).

Amendments Issued Since Publication

Amendment No.	Date of Issue	Text Affected

BUREAU OF INDIAN STANDARDS

Headquarters:

Manak Bhavan, 9 Bahadur Shah Zafar Marg, New Delhi 110002
Telephones: 2323 0131, 2323 3375, 2323 9402 *Website*: www.bis.gov.in

Regional Offices:		*Telephones*
Central	: Manak Bhavan, 9 Bahadur Shah Zafar Marg, NEW DELHI 110002	2323 7617, 2323 3841
Eastern	: 1/14, C.I.T. Scheme VII M, V.I.P. Road, Kankurgachi, KOLKATA 700054	2337 8499, 2337 8561, 2337 8626, 2337 9120
Northern	: Plot No. 4-A, Sector 27-B, Madhya Marg, CHANDIGARH 160019	26 50206, 265 0290
Southern	: C.I.T. Campus, IV Cross Road, CHENNAI 600113	2254 1216, 2254 1442, 2254 2519, 2254 2315
Western	: Manakalaya, E9 MIDC, Marol, Andheri (East), MUMBAI 400093	2832 9295, 2832 7858, 2832 7891, 2832 7892

Branches : AHMEDABAD. BENGALURU. BHOPAL. BHUBANESWAR. COIMBATORE. DEHRADUN. DURGAPUR. FARIDABAD. GHAZIABAD. GUWAHATI. HYDERABAD. JAIPUR. JAMMU. JAMSHEDPUR. KOCHI. LUCKNOW. NAGPUR. PARWANOO. PATNA. PUNE. RAIPUR. RAJKOT. VISAKHAPATNAM.

Published by BIS, New Delhi