ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Recommendation

ITU-T F.748.25 (09/2023)

SERIES F: Non-telephone telecommunication services

Multimedia services

Requirements for speech interaction of intelligent customer services

ITU-T F-SERIES RECOMMENDATIONS

Non-telephone telecommunication services

TELEGRAPH SERVICE	F.1-F.109
Operating methods for the international public telegram service	F.1-F.19
The gentex network	F.20-F.29
Message switching	F.30-F.39
The international telemessage service	F.40-F.58
The international telex service	F.59-F.89
Statistics and publications on international telegraph services	F.90-F.99
Scheduled and leased communication services	F.100-F.104
Phototelegraph service	F.105-F.109
MOBILE SERVICE	F.110-F.159
Mobile services and multidestination satellite services	F.110-F.159
TELEMATIC SERVICES	F.160-F.399
Public facsimile service	F.160-F.199
Teletex service	F.200-F.299
Videotex service	F.300-F.349
General provisions for telematic services	F.350-F.399
MESSAGE HANDLING SERVICES	F.400-F.499
DIRECTORY SERVICES	F.500-F.549
DOCUMENT COMMUNICATION	F.550-F.599
Document communication	F.550-F.579
Programming communication interfaces	F.580-F.599
DATA TRANSMISSION SERVICES	F.600-F.699
MULTIMEDIA SERVICES	**F.700-F.799**
ISDN SERVICES	F.800-F.849
UNIVERSAL PERSONAL TELECOMMUNICATION	F.850-F.899
ACCESSIBILITY AND HUMAN FACTORS	F.900-F.999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T F.748.25

Requirements for speech interaction of intelligent customer services

Summary

Recommendation ITU-T F.748.25 describes scenarios, high-level architecture, functional requirements, and performance requirements for speech interaction of intelligent customer services. Some detailed use cases and reference process for the creation of the knowledge base are described in Appendices I and II, respectively.

History *

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T F.748.25	2023-09-13	16	11.1002/1000/15622

Keywords

Artificial intelligence, customer service, requirements, speech interaction, use cases.

* To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T F.748.25

Requirements for speech interaction of intelligent customer services

1 Scope

This Recommendation specifies high-level architecture and requirements for speech interaction of intelligent customer services.

This Recommendation covers the following:

– High-level architecture of speech interaction of intelligent customer service;

– Functional requirements for speech interaction of intelligent customer service;

– Performance requirements for speech interaction of intelligent customer service.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

None.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 application [b-ITU-T Y.2091]: A structured set of capabilities, which provide value-added functionality supported by one or more services, which may be supported by an API interface.

3.1.2 artificial intelligence (AI) [b-ITU-T F.749.13]: An interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning.

3.1.3 automatic speech recognition (ASR) [b-ITU-T H.625]: A system that can recognize continuous speech, often having phoneme-sized references, using lexical, syntactic, semantic, and pragmatic knowledge, and reacts appropriately (therefore having interpreted the message and found the corresponding action to be taken).

3.1.4 knowledge base [b-ITU-T F.746.3]: A collection of knowledge resources that consist of structured and unstructured data. The knowledge base is used to provide information to the various applications that are related to information provision such as the QA system and search system.

3.1.5 knowledge graph [b-IEEE 2807]: Assemblies of knowledge elements and their relations described in a structured form.

3.1.6 mean opinion score (MOS) [b-ITU-T P.800]: The mean of opinion scores, i.e., of the values on a predefined scale that subjects assign to their opinion of the performance of the telephone transmission system used either for conversation or for listening to spoken material.

3.1.7 natural language processing (NLP) [b-ITU-T F.746.3]: A method that analyses text in natural languages through several processes such as part-of-speech recognition, syntactic analysis, and semantic analysis.

3.1.8 text-to-speech synthesis (TTS) [b-ITU-T P.10/G.100-Amendment 1]: A TTS process generates a speech signal from text codes. It is usually composed of two parts:

– a language-dependent text processing part (the high-level processing part), which generates from the character string (by reading rules, vocabulary and semantic analysis) a set of phonetic, prosodic, etc., parameters that are used by:

– an acoustical signal generating part, the synthesizer itself, which generates the audible speech.

3.1.9 voice recognition [b-ISO 9241-154]: Capability of a system to identify a specific person's voice.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 intelligent customer service: Customer service with intelligent capabilities, which is composed of some or all of the artificial intelligence technology capabilities such as speech recognition, natural language processing, intelligent search, and text-to-speech synthesis.

3.2.2 multi-factor authentication: Authentication using two or more factors to achieve authentication.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI Artificial Intelligence

ASR Automatic Speech Recognition

CRM Customer Relationship Management

IVR Interactive Voice Response

KB Knowledge Base

KG Knowledge Graph

MOS Mean Opinion Score

NLP Natural Language Processing

TTS Text-to-Speech

5 Conventions

In this Recommendation:

The keywords "be required to" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

The keywords "can optionally" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option and the feature can be optionally enabled by the network

operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with the specification.

6 Overview

Intelligent customer service is based on big data and artificial intelligence (AI) technologies such as natural language processing (NLP), machine learning, and large-scale knowledge management. It enables communication between enterprises and users, thereby improving operational efficiency and reducing costs.

Figure 1 shows the speech interaction scenarios of intelligent customer service. These include four major scenarios.

– Voice navigation scenario

Voice navigation scenario mainly describes the scenarios in which the user speaks the required service through the communication terminal, and the intelligent customer service system interacts with the customer and can quickly and accurately find relevant content for the user and make a voice reply. If the user needs to know more information, the intelligent customer service system can directly transfer the voice stream to the seat, which can provide more detailed services.

– Intelligent outbound scenario

The intelligent outbound scenario mainly describes the scenarios that communicate with customers according to the preset outbound logic. Through multiple rounds of conversation and vocabulary guidance, it can complete the business function of outbound calls.

– Operator assistant scenario

Operator assistant scenario mainly describes the scenarios which can obtain customer information through customer relationship management (CRM) and gives prompts to operators on the computer screen in real time to help operators communicate with the customers.

– Service quality inspection scenario

The service quality inspection scenario mainly describes which system performs online analysis of the voice stream obtained by the call centre in real time through the communication network, or which system downloads the customer service audio recording file for offline analysis and processing to generate the corresponding quality inspection analysis results.

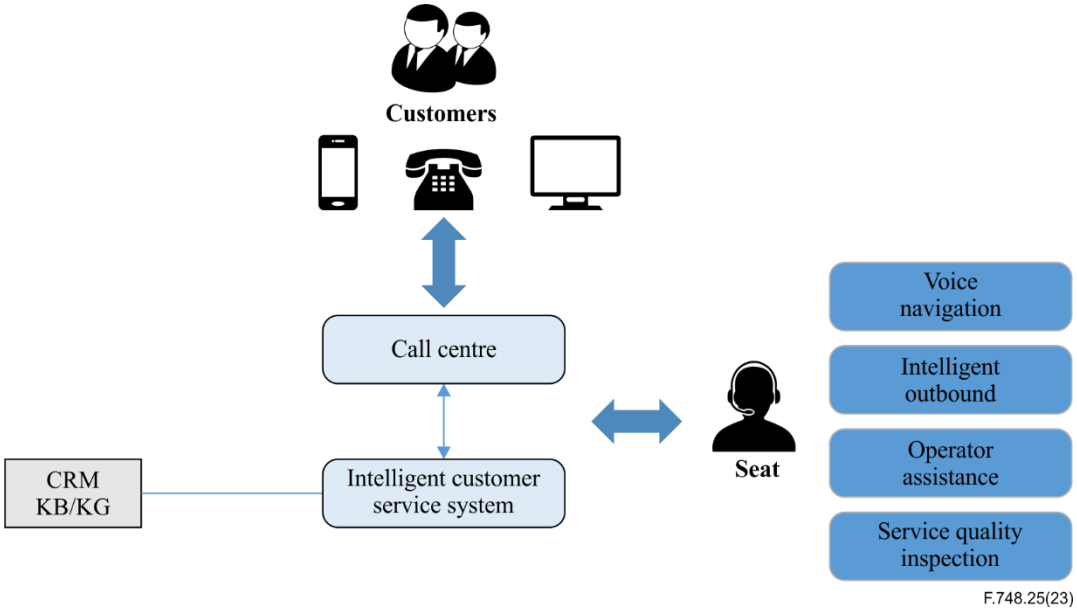

Figure 1 – Speech interaction scenarios in intelligent customer services

7 High-level architecture

Figure 2 presents the high-level architecture for speech interaction of intelligent customer services according to the use cases.

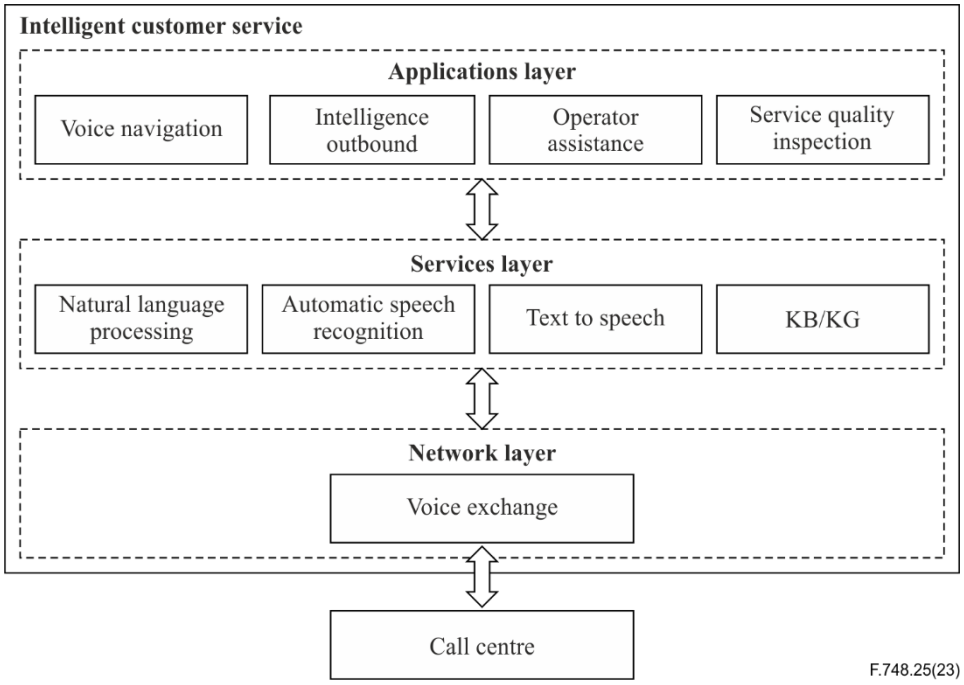

Figure 2 – High-level architecture for speech interaction of intelligent customer services

The high-level architecture for speech interaction of intelligent customer services covers three layers: applications, services and networks. A general description of these three blocks is presented below.

– **Applications layer**: Applications layer is the software for realizing the speech interaction of intelligent customer service. The applications include voice navigation, intelligence outbound, operator assistance and service quality inspection, etc.

  A detailed description of the applications can be found in Appendix I.

– **Services layer**: Services layer is the basic AI functional modules for supporting applications. The services include natural language processing (NLP), automatic speech recognition (ASR), text-to-speech synthesis (TTS), and knowledge base (KB)/knowledge graph (KG).

  The NLP function uses algorithms to understand and manipulate human language. The NLP function can realize applications of intelligent customer service through algorithms such as intention understanding, named entity recognition, text classification, etc.

  The ASR function uses algorithms to convert human speech into text. The ASR function can realize multiple applications in intelligent customer service through single-sentence speech recognition, continuous speech recognition, etc.

  The TTS function uses algorithms to synthesize natural-sounding speech in multiple languages. This function can be used in user interaction applications of intelligent customer service.

  The KB/KG contains information about words, sentences, and the relationship between knowledge. It describes all the knowledge information and the relationship between each piece of knowledge information used in the intelligent customer service.

– **Network layer**: Network layer is the communication network's capability for voice exchange between the call centre and services layer. The voice exchange supports multimedia voice, voice call in and call out, voice transfer and voice recording functions, etc.

8 Requirements

8.1 Functional requirements

The functional requirements for intelligent customer service are as follows:

– The speech interaction of intelligent customer service is required to have the ability to recognize keywords in sentences;

– The speech interaction of intelligent customer service is required to support speech interaction with multiple access channels, and supports one or more speech interaction methods such as fixed telephone network, mobile telephone network, Internet, and other access methods;

– The presentation of the speech interaction results of the intelligent customer service is required to include the results of TTS, and can optionally include the presentation of information such as text and graphics;

– The speech interaction of intelligent customer service is recommended to support automatic identification of voice start point and end point;

– The speech interaction of intelligent customer service can optionally support the dynamically generate guidance and prompt words based on the results of ASR and NLP;

– The speech interaction of intelligent customer service can optionally support the service quality inspection of operators based on the results of ASR and NLP;

– The ASR function is recommended to have the ability to recognize continuous speech;

– The ASR function is recommended to support the language recognition function, that is, to determine the language of the voice based on the user's voice input;

– The ASR function is recommended to support mixed language recognition, that is, support the recognition of different language words in sentences;

– The NLP function of intelligent customer service is recommended to support the understanding of abbreviations, aliases, locations and numbers;

– The NLP function of intelligent customer service is recommended to support the fault-tolerant understanding of typical misleading words;

– The NLP function of intelligent customer service is recommended to support multiple rounds of dialogue;

– The NLP function of intelligent customer service can optionally support the understanding of subordinate relationships, the understanding of partial-overall relationships, the understanding of geographic relationships, the understanding of causality, the understanding of sequential relationships, and the understanding of proximal relationships;

– The TTS function of intelligent customer service is required to support multi-language synthesis and multi-timbre synthesis such as young female voices and young male voices;

– The TTS function of intelligent customer service can optionally support mixed language and personalized TTS;

– The purpose of building a KB of intelligent customer service is mainly to organize and manage knowledge according to the business requirements, including keyword libraries, service processes, question templates, or KGs constructed in the form of triples.

a) The KB of intelligent customer service is required to have the function of keyword identification and maintenance according to the business requirements and support the addition, deletion, and modification of characteristic keywords of customer service system services.

b) The KB of intelligent customer service is required to have the maintenance function of static prompts, that is, the guiding prompt function according to the business

requirements, and support the addition, deletion, and modification of static prompting functions of the customer service system services.

c) The KB of intelligent customer service is required to have the service process maintenance function according to the business requirements, and support the addition, deletion, and modification of the service process of the customer service system.

d) The KB of intelligent customer service is required to have the function of editing question templates according to the business requirements, and support the addition, deletion, and modification of semantic expressions of customer service questions.

e) The KB of intelligent customer service can optionally support the knowledge construction based on the KG technology in the form of ontology/triple.

– The speech interaction of intelligent customer service that will use voice recognition for authentication, or verification for any use case, especially for critical transactions like services provisioning, financial transactions, or dealing with personal or financial data, is required to consider the implementation of multi-factor authentication beyond voice recognition. This is to protect customers' sensitive information and accounts from voice cloning trials by fraudsters by creating synthetic audio recordings of a person's voice that are often indistinguishable from authentic voices;

– The speech interaction of intelligent customer service is recommended to have the below measures to detect and prevent fraudulent activities such as suspicious account activity, unusual recognition attempts, or transaction patterns during the user verification process:

a) The system can optionally monitor transactions in real-time to detect any suspicious activity. While machine learning algorithms can optionally be used to analyse patterns of fraudulent behaviour and identify potential fraudsters.

b) The system can optionally send alerts to designated personnel when it detects suspicious activity.

c) The system can optionally block transactions that have a high-risk score to prevent fraud.

d) The system can optionally analyse user behaviour to detect any anomalies or suspicious activity.

8.2 Non-functional requirements

8.2.1 Performance requirements

8.2.1.1 Response time of speech interaction

The response time of speech interaction refers to the time between the end of the user input and the return of the result.

The interactive response time calculation method is shown in the formula.

$$T_{ack} = t_r - t_e$$

T_{ack} – Response time

t_r – Result output moment

t_e – Voice input end time

If the speech interaction system supports the recognition result to be returned multiple times, t_r should be the time when the first part of the recognition result is returned.

In the case of multiple rounds of speech interaction process, T_{ack} refers to the response time of each round of speech interaction.

In order to ensure the service quality of intelligent customer service, T_{ack} is recommended to be less than an appropriate threshold.

8.2.1.2 Speech interaction success rate

In the interaction process of the speech interaction system, the user can obtain the result within the predetermined interaction rounds, and it can then be judged that the current rounds of interaction are successful; otherwise, it is judged that these rounds of interaction have failed.

The calculation method of the interaction success rate is as follows:

$$P_s = \frac{S}{S+F} \times 100\%$$

P_s – Interaction success rate

S – The number of successful interactions

F – The number of failed interactions

In order to ensure the service quality of intelligent customer service, P_s is recommended to be higher than an appropriate threshold.

8.2.1.3 ASR accuracy index

In the speech interaction system, it is recommended that the word accuracy of keyword speech recognition should be calculated in different noise environments (e.g., noise intensity below 50 dB and noise intensity of 60 dB to 65 dB, etc.).

In the speech interaction system, it is recommended that the word accuracy rate of continuous speech recognition should be calculated in, in different noise environments (e.g., noise intensity below 50 dB and noise intensity of 60 dB to 65 dB, etc.).

In order to ensure the service quality of intelligent customer service, it is recommended that the ASR accuracy should be higher than an appropriate threshold.

8.2.1.4 TTS evaluation

TTS test adopts the mean opinion score (MOS) test method for evaluation. Through the result of the intelligent customer service speech synthesis, it tests the difference in sound quality, intelligibility and naturalness between the synthesized speech and the human speech, numerically quantifies the test result with the average MOS score, and then records the average result.

In order to ensure the service quality of intelligent customer service, MOS is recommended to be greater than or equal to an appropriate threshold.

8.2.1.5 NLP accuracy index

The NLP ability is recommended to be tested by various people with multiple rounds. The calculation method for the correct rate of NLP is as follows.

$$R_{SS} = \frac{N_{SS}}{N} \times 100\%$$

R_{ss} – NLP rate

N_{ss} – The number of times the intention and semantic elements are correctly judged

N – The total number of correct texts

In order to ensure the service quality of intelligent customer service, R_{ss} is recommended to be higher than an appropriate threshold.

Appendix I

Use cases of intelligent customer services

(This appendix does not form an integral part of this Recommendation.)

I.1 Use cases

I.1.1 Use cases for full speech interaction of intelligent customer services

The full speech interaction use case means that all the functions are completed by the speech interaction system, and there is no operator participation during the interaction. It may include:

a) The call centre recognizes the customer's voice input through ASR and NLP technology, then automatically responds to the customer through TTS technology.

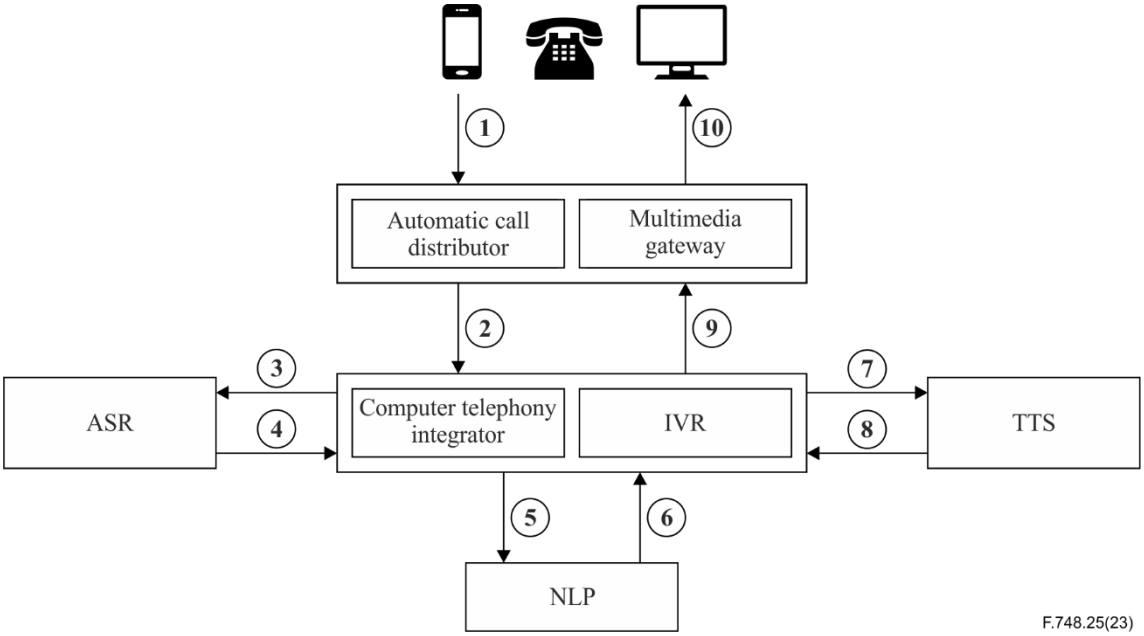

Figure I.1 – Example of full speech interaction of intelligent customer service (Model A)

The description of the workflow is as below:

1) The customer calls the call centre using a cell phone, fixed telephone or a computer and speaks according to voice prompts;
2) The automatic call distributor / multimedia gateway of the call centre transfers the speech to the computer telephony integrator / interactive voice response (IVR);
3) IVR transfers the call to the ASR function;
4) The ASR function converts the speech to text and returns the result to the IVR;
5) IVR uses NLP technology to recognize user intent for the text;
6) NLP function returns the result to the IVR;
7) IVR matches the predefined process based on the results of NLP to synthesize the next speech;
8) TTS function returns the speech to the IVR;
9) IVR transfers the speech to the automatic call distributor / multimedia gateway through the call connection;

10) The automatic call distributor / multimedia gateway transfers the speech to the customer. Then the customer starts the next round of dialogue.

b) Instead of traditional manual telephone return visits and telephone notifications, the standard services of the call centre can be implemented through a dialogue between the robot and the user.

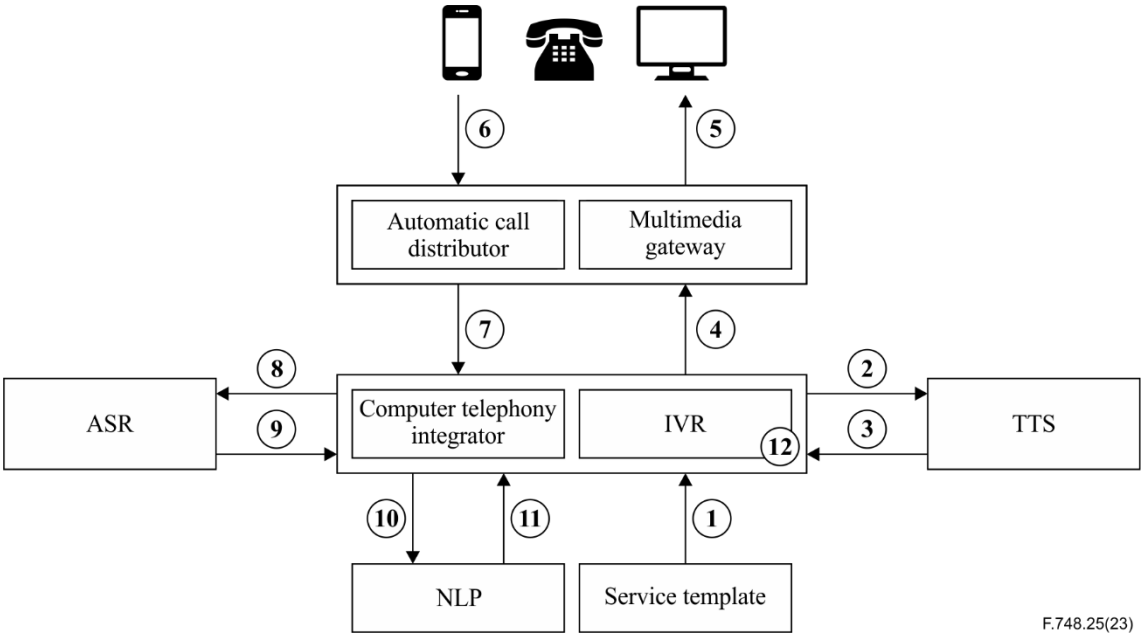

Figure I.2 – Example of full speech interaction of intelligent customer service (Model B)

The description of the workflow is as below:

1) The service template is pre-installed in the IVR;
2) The call centre initiates a call connection and the IVR requests the TTS function for the service template;
3) The TTS function returns the speech to the IVR;
4) IVR transfers the speech to the automatic call distributor / multimedia gateway through the call connection;
5) The automatic call distributor / multimedia gateway transfers the speech to the customer;
6) The customer answers the call and speaks according to voice prompts;
7) The automatic call distributor / multimedia gateway of the call centre transfers the speech to the computer telephony integrator / IVR;
8) IVR transfers the call to the ASR function;
9) The ASR function converts the speech to text and returns the result to the IVR;
10) IVR uses NLP technology to recognize user intent for the text;
11) NLP function returns the result to IVR;
12) IVR matches the predefined template based on the results of the NLP to the next round of dialogue.

I.1.2 Use cases for partial speech interaction of intelligent customer services

Partial speech interaction use cases refers to the participation of the operator throughout the speech interaction process. It may include:

a) During the speech interaction process, the user cannot obtain the target result, or the target result does not conform to the service specification. The system transfers the interactive session to the operator, and the service will be provided by the operator.

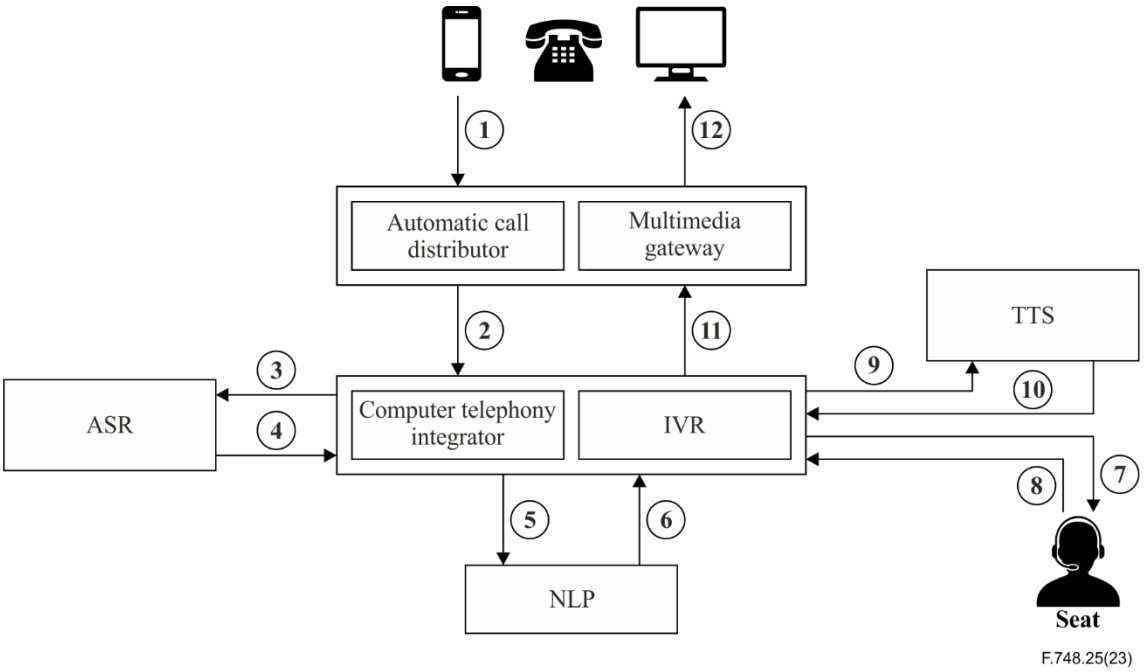

Figure I.3 – Example of partial speech interaction of intelligent customer service (Mode A)

The description of the workflow is as below:

1) The customer calls the call centre using a cell phone, fixed telephone or computer and speaks according to voice prompts;
2) The automatic call distributor / multimedia gateway of the call centre transfers the speech to the computer telephony integrator / IVR;
3) IVR transfers the call to the ASR function;
4) The ASR function converts the speech to text and returns the text to the IVR;
5) IVR uses NLP technology to recognize user intent for the text;
6) The NLP function returns the result to the IVR;
7) IVR cannot match the right process based on the results of NLP and transfers the call to the seat;
8) The operator communicates with the customer and helps the customer deal with problems;
9) The operator determines to return the result with the IVR and transfers the call to the IVR. IVR requests the TTS function to synthesize the speech;
10) The TTS function returns the speech to IVR;
11) IVR transfers the speech to the automatic call distributor / multimedia gateway through the call connection;
12) The automatic call distributor / multimedia gateway transfers the speech to the customer.

b) Speech interaction system only performs certain functions, such as IVR navigation. That is, the user speaks the specific words, and after the ASR, the speech is transferred to a specific operator for follow-up service.

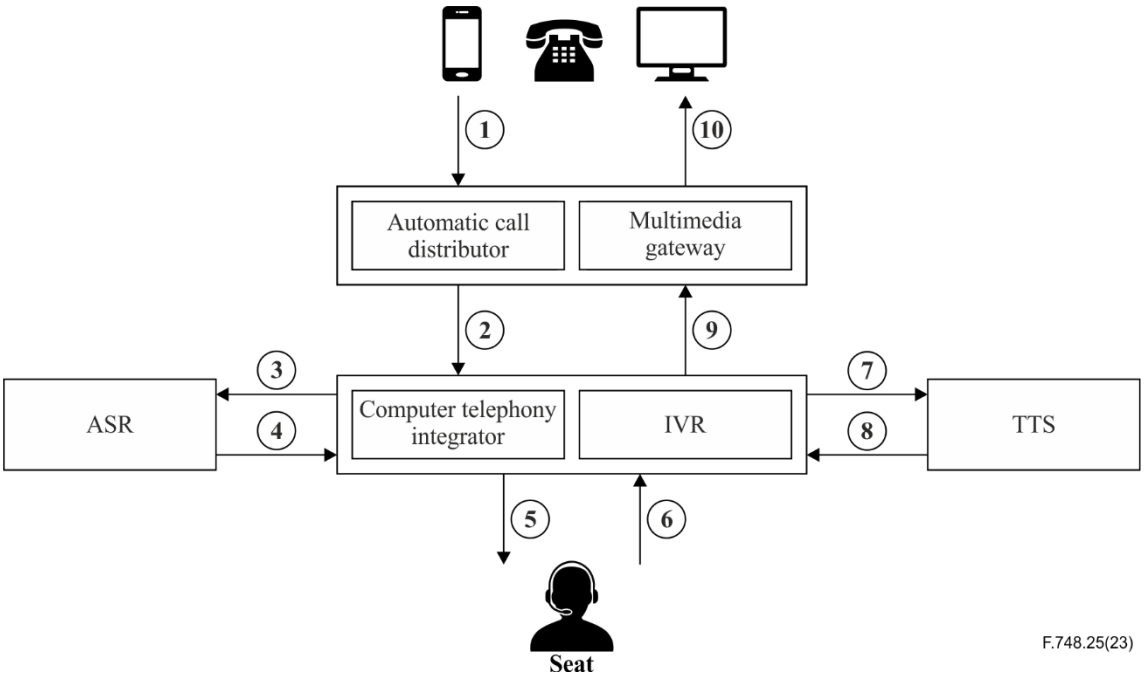

Figure I.4 – Example of partial speech interaction of intelligent customer service (Mode B)

The description of the workflow is as below:

1) The customer calls the call centre using a cell phone, fixed telephone or computer and speaks according to voice prompts;
2) The automatic call distributor / multimedia gateway of the call centre transfers the speech to the computer telephony integrator / IVR;
3) IVR transfers the call to the ASR function;
4) The ASR function converts the speech to text and returns the result to the IVR;
5) IVR matches the right process based on the results of the ASR and transfers the call to the specific seat;
6) The operator communicates with the customer and helps the customer deal with the problems;
7) The operator determines to return the result with IVR and transfers the call to the IVR. IVR requests the TTS function to synthesize the speech;
8) TTS function returns the speech to the IVR;
9) IVR transfers the speech to the automatic call distributor / multimedia gateway through the call connection;
10) The automatic call distributor / multimedia gateway transfers the speech to the customer.

I.1.3 Use cases for service quality inspection of intelligent customer services

Traditional service quality inspection requires a lot of workloads. Through ASR and NLP techniques, the operator's service quality can be automatically evaluated.

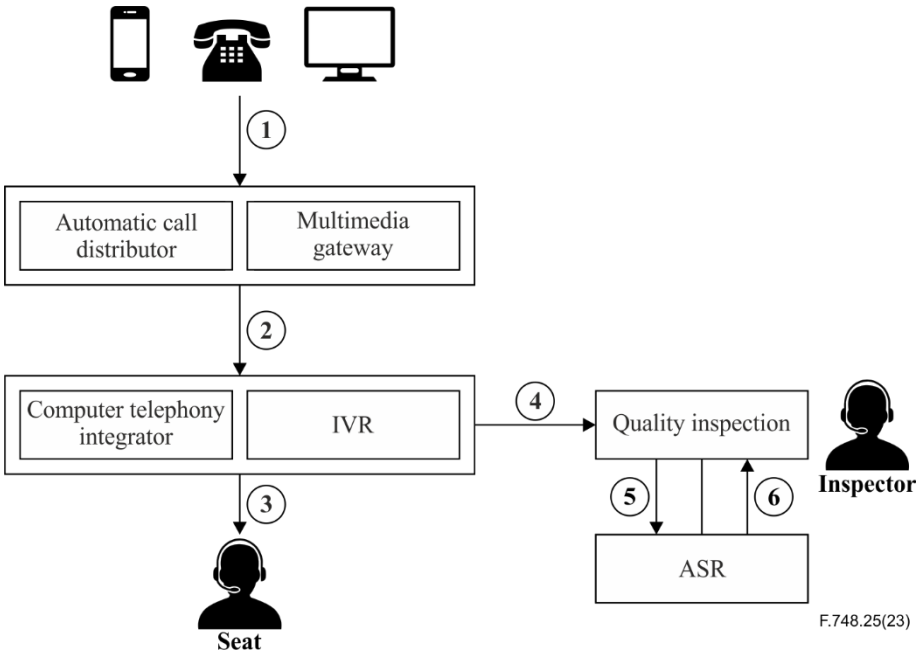

Figure I.5 – Example of service quality inspection of intelligent customer services

The description of the workflow is as below:

1) The customer calls the call centre using a cell phone, fixed telephone or computer and speaks according to voice prompts;
2) The automatic call distributor / multimedia gateway of the call centre transfers the speech to the computer telephony integrator / IVR;
3) The operator communicates with the customer and helps the customer deal with the problems;
4) The quality inspection function can perform quality inspection online/offline according to the voice of the operator;
5) The quality inspection function transfers the voice to the ASR function;
6) The quality inspection module supports voice information detection such as speech speed and mute in different roles during the conversation.

I.1.4 Use cases for operator assistance of intelligent customer services

The operator assistance system can identify the conversation between the operators and the customers through ASR technology during the call and capture the key information through the NLP technology. It can then obtain customer information through the CRM and give prompts to operators on the computer screen in real time (including knowledge points and verbal reminders, automated extraction of customer portraits, analysis of customer purchase intentions, etc.) to help operators communicate with the customers.

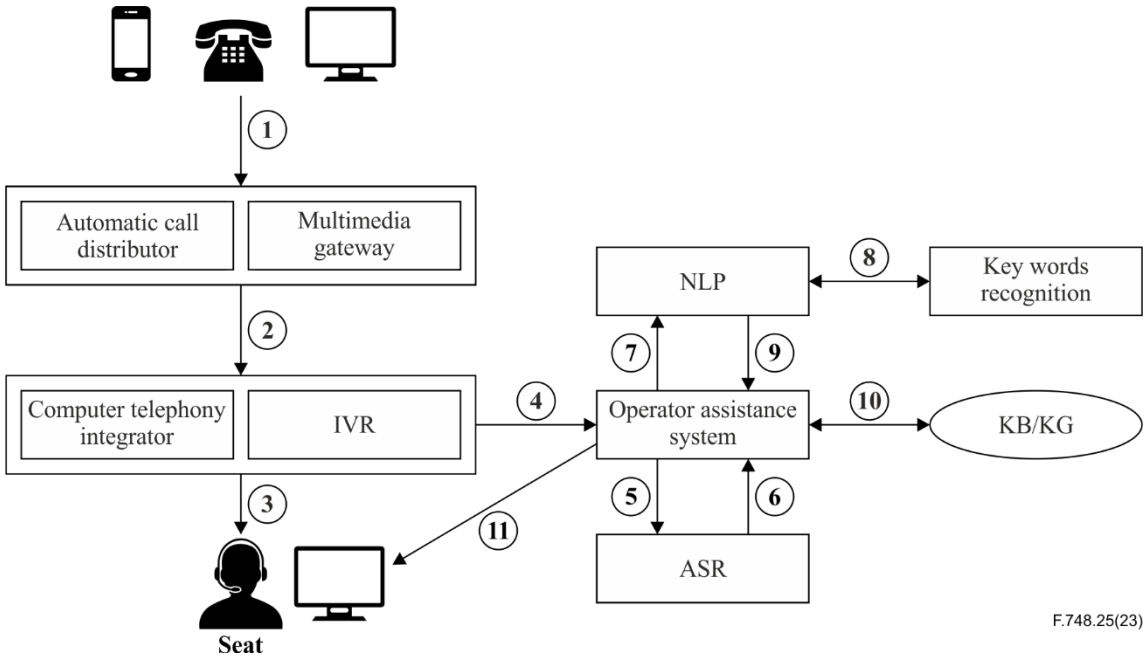

Figure I.6 – Example of operator assistance of intelligent customer services

The description of the workflow is as below:

1) The customer calls the call centre using a cell phone, fixed telephone or computer and speaks according to voice prompts;
2) The automatic call distributor / multimedia gateway of the call centre transfers the speech to the computer telephony integrator / IVR;
3) The operator communicates with the customer and helps the customer deal with the problems;
4) The operator assistance system can give the information to the operator according to the voice of the operator and the customer;
5) The operator assistance system transfers the voice to the ASR function;
6) The ASR function converts the speech to text and returns the text to the operator assistance system;
7) The operator assistance system uses the NLP technology to perform NLP for the text;
8) Key words recognition module performs keywords' detection;
9) The NLP function returns the text and keywords result to the operator assistance system;
10) The operator assistance system obtains customer information from KB/KG;
11) The operator assistance system gives prompts to operators on the computer screen in real time.

I.1.5 Use case of a phone-based voice authentication

Use case of a phone-based voice authentication, as an example of using voice recognition within an intelligent customer service.

The below workflow outlines a phone-based voice authentication that uses a combination of identifying information, voice input, and additional authentication factors to granting access to users' accounts or other protected resources.

The description of the workflow is as below:

a) The user calls the authentication system through IVR using their registered phone number.

b) The system prompts the user to provide their username or other identifying information such as an account number or national identification number.

c) The user provides their username or other identifying information using a voice input.

d) The system verifies the user based on the provided information and prompts the user to provide a voice sample for verification.

e) The user provides a voice sample for verification.

f) If the voice sample matches, the system prompts the user to provide additional authentication factors such as password, or PIN, and a physical security token.

g) The user provides the additional authentication factor(s) as requested.

h) If all authentication factors are successfully verified, the system grants access to the user's account or other protected resource.

It is worth noting that this is just one example workflow, and the actual implementation may vary depending on the specific requirements and technologies used.

Appendix II

Reference process for the creation of the knowledge base

(This appendix does not form an integral part of this Recommendation.)

II.1 Overview

The construction of the KB of the intelligent customer service system starts with a single domain, and then expands to multiple domains.

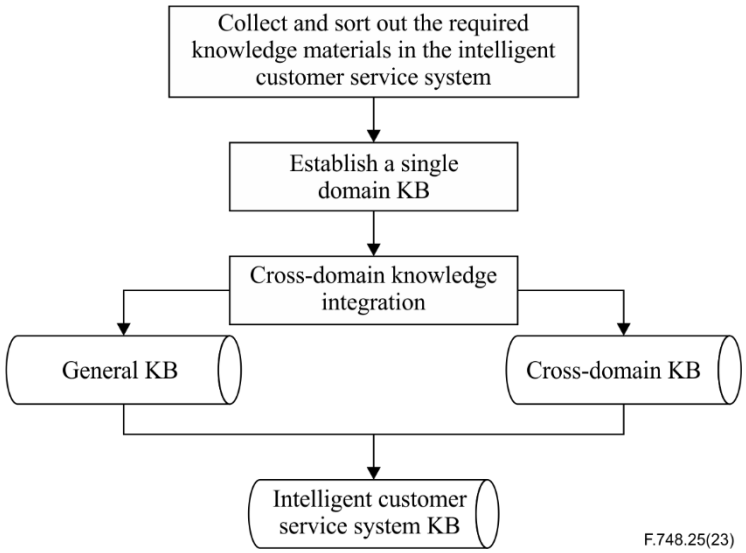

Figure II.1 – Basic process of a KB establishment

Figure II.1 describes a basic process example for a knowledge base (KB) creation. First, a single-domain KB is established through natural language technologies such as entity recognition and relation extraction. Then, through knowledge mapping and fusion technologies, cross-domain knowledge integration is realized, and then the general domain KB and cross-domain KB are established.

II.2 Single domain KB establishment

The intelligent customer service system has accumulated a large number of fuzzy and inaccurate knowledge in the process of use. The existence of these knowledge points is a large amount of incomplete, noisy, fuzzy, random practical application data (involving natural language technology). Therefore, the process of extracting hidden but potentially useful information and knowledge is very important and necessary, so that much historical data that could not be understood before can be used.

The process of single domain knowledge mining is divided into the following steps:

– Keyword libraries construction: Based on the NLP technology, discovering concepts in the domain through entity recognition and syntactic analysis technologies, mining the relationships between concepts, and automatically or semi-automatically constructing domain semantic lexicons based on semantic network technology;

– Construction of a candidate object: Semantic understanding and induction of concepts in the semantic lexicon, and construction of a candidate object;

– Submit for manual review: This semi-automatic knowledge mining method effectively speeds up the process of domain KB construction and reduces costs.

II.3 Cross-domain KB establishment

Based on a KB of a single domain, knowledge of a new domain is gradually integrated. Make the KB adjust with the integration of new domain knowledge, adapt it to the requirements of the corresponding multiple domains, and finally realize the integration of a cross-domain knowledge.

– Integration of words in various fields: Merge common words that overlap in the KB of various fields and classify professional words in vertical fields according to different fields.

– Combination of object classes: To integrate knowledge of the new domain into the existing KB. Firstly, it is necessary to find similar object classes in the existing knowledge by the method of object class matching, and then merge. Object class merging is to integrate multiple object classes together, eliminate overlapping parts and remove inconsistent parts, so that the new object class is more reasonable and can better express the concept of these merged object classes.

– Combination of knowledge: In addition to the specific professional knowledge content of each domain, there is also a part of knowledge that is common to multiple domains. The general knowledge content in various domains needs to be extracted to form a general KB.

Bibliography

[b-ITU-T F.746.3] Recommendation ITU-T F.746.3 (2015), *Intelligent question answering service framework.*

[b-ITU-T F.749.13] Recommendation ITU-T F.749.13 (2021), *Framework and requirements for civilian unmanned aerial vehicle flight control using artificial intelligence.*

[b-ITU-T H.625] Recommendation ITU-T H.625 (2017), *Architecture for network-based speech-to-speech translation services.*

[b-ITU-T P.10/G.100] Recommendation ITU-T P.10/G.100 (2017) – Amendment 1 (2019), *New definitions for inclusion in Recommendation ITU-T P.10/G.100. Vocabulary for performance, quality of service and quality of experience Amendment 1.*

[b-ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality.*

[b-ITU-T Y.2091] Recommendation ITU-T Y.2091 (2011), *Terms and definitions for next generation networks.*

[b-IEEE 2807] IEEE 2807-2022, *IEEE Standard for Framework of Knowledge Graphs.* <https://standards.ieee.org/ieee/2807/7525/>

[b-ISO 9241-154] ISO 9241-154:2013, *Ergonomics of human-system interaction – Part 154: Interactive voice response (IVR) applications.* <https://www.iso.org/standard/52172.html#:~:text=ISO%209241%2D154%3A2013%20gives,ASR)%20as%20the%20input%20mechanism.>

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	**Non-telephone telecommunication services**
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Published in Switzerland
Geneva, 2023