

International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

H.862.3

(08/2020)

SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS

E-health multimedia systems, services and applications –
Multimedia e-health data exchange services

**Requirements of the voice management
interface for human-care services**

Recommendation ITU-T H.862.3

ITU-T



ITU-T H-SERIES RECOMMENDATIONS
AUDIOVISUAL AND MULTIMEDIA SYSTEMS

CHARACTERISTICS OF VISUAL TELEPHONE SYSTEMS	H.100–H.199
INFRASTRUCTURE OF AUDIOVISUAL SERVICES	
General	H.200–H.219
Transmission multiplexing and synchronization	H.220–H.229
Systems aspects	H.230–H.239
Communication procedures	H.240–H.259
Coding of moving video	H.260–H.279
Related systems aspects	H.280–H.299
Systems and terminal equipment for audiovisual services	H.300–H.349
Directory services architecture for audiovisual and multimedia services	H.350–H.359
Quality of service architecture for audiovisual and multimedia services	H.360–H.369
Telepresence, immersive environments, virtual and extended reality	H.420–H.439
Supplementary services for multimedia	H.450–H.499
MOBILITY AND COLLABORATION PROCEDURES	
Overview of Mobility and Collaboration, definitions, protocols and procedures	H.500–H.509
Mobility for H-Series multimedia systems and services	H.510–H.519
Mobile multimedia collaboration applications and services	H.520–H.529
Security for mobile multimedia systems and services	H.530–H.539
Security for mobile multimedia collaboration applications and services	H.540–H.549
VEHICULAR GATEWAYS AND INTELLIGENT TRANSPORTATION SYSTEMS (ITS)	
Architecture for vehicular gateways	H.550–H.559
Vehicular gateway interfaces	H.560–H.569
BROADBAND, TRIPLE-PLAY AND ADVANCED MULTIMEDIA SERVICES	
Broadband multimedia services over VDSL	H.610–H.619
Advanced multimedia services and applications	H.620–H.629
Ubiquitous sensor network applications and Internet of Things	H.640–H.649
IPTV MULTIMEDIA SERVICES AND APPLICATIONS FOR IPTV	
General aspects	H.700–H.719
IPTV terminal devices	H.720–H.729
IPTV middleware	H.730–H.739
IPTV application event handling	H.740–H.749
IPTV metadata	H.750–H.759
IPTV multimedia application frameworks	H.760–H.769
IPTV service discovery up to consumption	H.770–H.779
Digital Signage	H.780–H.789
E-HEALTH MULTIMEDIA SYSTEMS, SERVICES AND APPLICATIONS	
Personal health systems	H.810–H.819
Interoperability compliance testing of personal health systems (HRN, PAN, LAN, TAN and WAN)	H.820–H.859
Multimedia e-health data exchange services	H.860–H.869
Safe listening	H.870–H.879

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

Recommendation ITU-T H.862.3

Requirements of the voice management interface for human-care services

Summary

Recommendation ITU-T H.862.3 describes requirements of the voice management interface for human-care services. This Recommendation focuses on the classification of users and services for the voice user interface (UI) in.

Human-care services are related to the improvement of the quality of life by providing what is necessary for the health, welfare and protection of people. Various voice-based services are being developed to have nursing robots take care of patients and be able to identify problems based on conversations with the patients. This includes the development of technology for voice-based early diagnosis of dementia. Recent research and the use of speech recognition systems have improved the basic keyboard-based user interface with a natural dialogue user interface. In addition, it has become possible to manage the customer's health by, for example, checking and improving the customer's health status.

This Recommendation proposes a service model through the speech recognition interface in human-care services using several scenarios to apply speech recognition technology.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T H.862.3	2020-08-13	16	11.1002/1000/14354

Keywords

Speech, voice, voice user interface (VUI).

* To access the Recommendation, type the URL <http://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID. For example, <http://handle.itu.int/11.1002/1000/11830-en>.

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, 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 TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2020

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

	Page
1 Scope	1
2 References.....	1
3 Definitions	1
3.1 Terms defined elsewhere	1
3.2 Terms defined within this Recommendation.....	1
4 Abbreviations and acronyms	1
5 Conventions	2
6 Introduction	2
7 Service framework.....	3
7.1 Voice user interface	3
7.2 Speech recognition	4
7.3 Classification according to purpose of use	5
7.4 Functional entities	5
8 Requirements for the services.....	6
8.1 Considerations for service users	6
8.2 Consideration for service environments	6
8.3 Personalized services	6
9 Applications.....	7
9.1 Visually impaired user services	7
9.2 The voice UI application for healthcare services	7
9.3 Health improvement and promotion for elderly	8
Bibliography.....	9

Recommendation ITU-T H.862.3

Requirements of the voice management interface for human-care services

1 Scope

This Recommendation describes requirements of the voice management interface for human-care services. This Recommendation specifically describes the classification of users and services for the voice user interface (UI) in human-care services. The service requirements are defined, and the service applications are introduced.

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 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.

3.1.2 speech recognition [b-ITU-T H.703]: A kind of user interface for translation of spoken words into text.

3.2 Terms defined within this Recommendation

This Recommendation defines the following terms:

3.2.1 human-care services: Services provided to people to help them stabilize their life and find self-sufficiency through guidance, counselling, treatment and the provision of basic needs. It includes the delivery of healthcare.

3.2.2 voice user interface (VUI): Set of hardware and software components that enables determining the appropriate course of action following a person's voice commands and questions (via speech recognition) and responding to them.

3.2.3 speech to text: The process of converting speech input into digital text.

3.2.4 text to speech: The process of converting digital text input into speech.

3.2.5 voice trigger: A system- or user-defined voice command that prompts a voice user interface (VUI) to start voice processing.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI	Artificial Intelligence
AM	Acoustic Model
ASR	Automatic Speech Recognition
IoT	Internet of Things
LM	Language Model
NLP	Natural Language Processing
PHR	Personal Health Report
UI	User Interface
VUI	Voice User Interface
WER	Word Error Rate

5 Conventions

In this Recommendation:

- The keywords "is 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 needs not be present to claim conformance.
- The keywords "can optionally" and "may" indicate an optional requirement which is permissible, without implying any sense of being recommended. These terms are 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 Introduction

Human-care services are related to the improvement of the quality of life by providing what is necessary for the health, welfare, maintenance and protection of people. Various voice-based services are being developed to have nursing robots take care of patients and be able to identify problems based on conversations with the patients. This includes the development of technology for voice-based early diagnosis of dementia. This Recommendation proposes a service model through the speech recognition interface in human-care services using several scenarios to apply speech recognition technology.

Speech recognition has been used in the traditional healthcare sector only as a means to enhance the convenience of medical personnel in producing medical records. However, recent research and the use of speech recognition systems have improved the basic keyboard-based user interface with a natural dialogue user interface. In addition, it has become possible to manage the customer's health by, for example, checking and improving the customer's health status.

Consider the following aspects of speech recognition interfaces applicable to human-care services:

- Speech recognition interfaces can be applied for convenience, diagnosis and health improvement.
- Speech recognition interfaces should be categorized and serviced depending on the type of user, such as medical staff, healthy people, disabled people and patients.
- For application purposes, the speech recognition interface should be used to identify changes in the speaker's speech characteristics, as those can be used as a biomarker for certain medical conditions, e.g., geriatric syndromes, cognitive disorders, etc.

At a time when healthcare, speech recognition, and artificial intelligence (AI) technologies are widely proposed and used, the relevant standards may be an important opportunity to facilitate the development of the technology in the industry.

Healthcare service users need a type of interface different from the voice interface for general-purpose users. The voice of an elderly person is inaccurate and has a lower pitch than that of younger persons. In particular, users with certain diseases such as dementia have a less accurate pronunciation. An example of smart health services based on speech recognition interface is shown in Figure 1. In addition, ongoing management of users through the voice interface can be utilized as a biomarker for an early diagnosis of diseases based on changes in speech. To include these parts, a service should be developed that includes consideration of speech recognition, speech diagnosis and changes in speech and word usage. This recommendation presents a voice user interface model that includes these aspects.

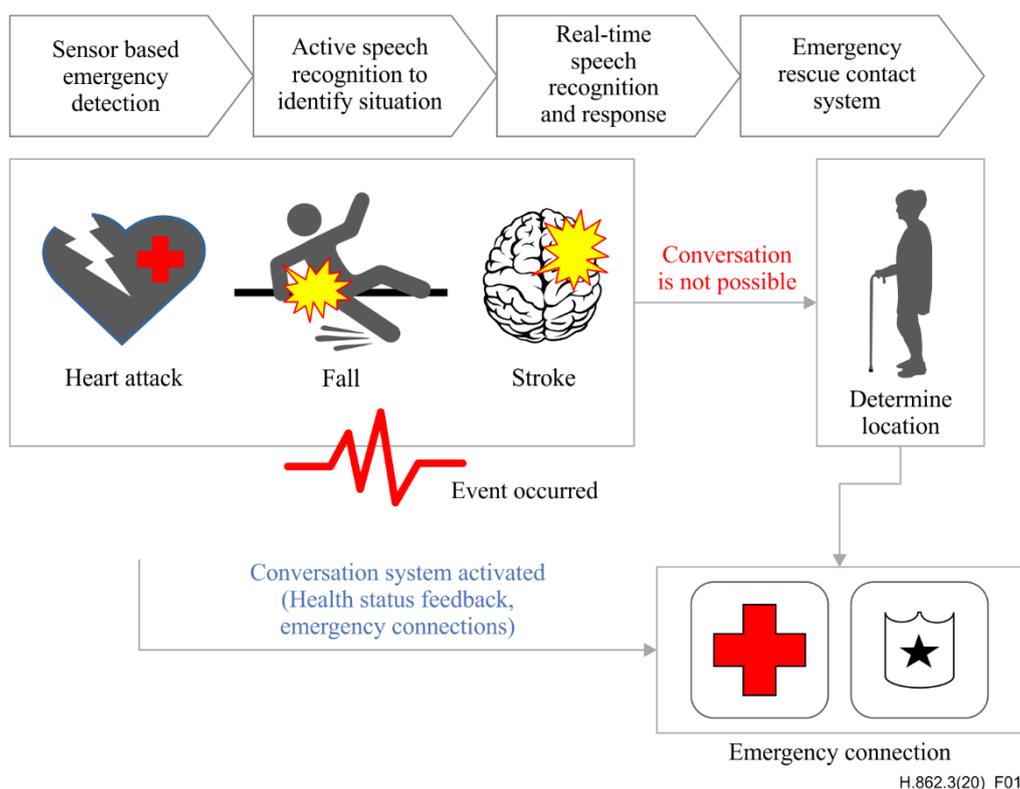


Figure 1 – Example of smart health services based on speech recognition interface

7 Service framework

7.1 Voice user interface

These clauses describe the general voice user interface. The speech recognition process assumes that normal conversation is possible. Further, it is assumed that there is no problem in the visual or auditory senses, and that there is no problem in confirming the feedback result. Figure 2 gives an overview of the general voice UI.

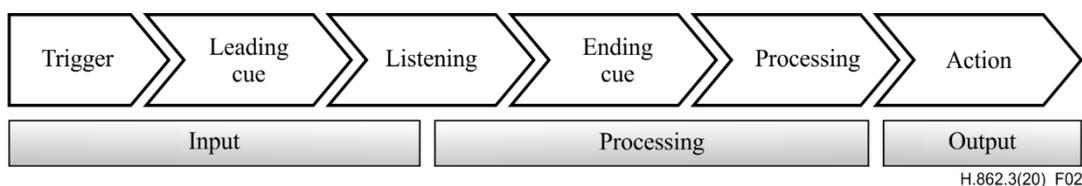
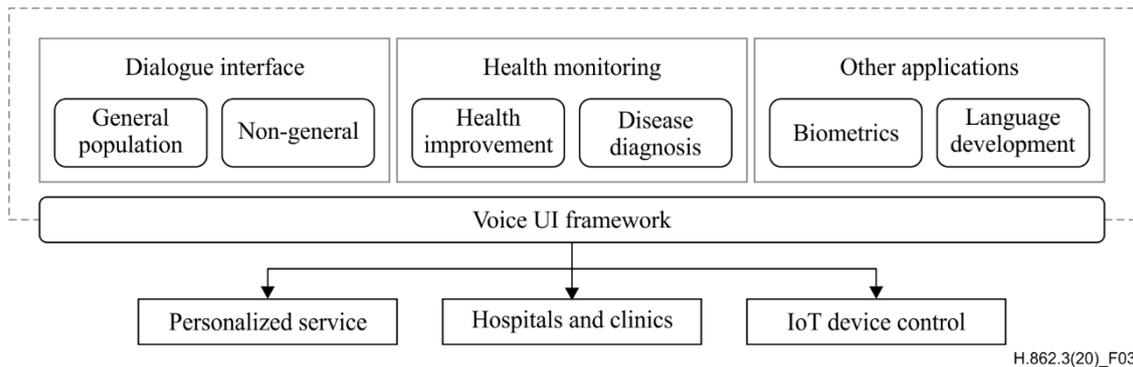


Figure 2 – General voice UI

In this Recommendation, speech recognition includes people who may have physical or mental impairments. In this sense, it describes the factors that should be considered in the speech recognition process in terms of service. An example of smart health services based on speech recognition interface is given in Figure 3.



H.862.3(20)_F03

Figure 3 – Example of smart health services based on speech recognition interface

7.2 Speech recognition

In speech recognition, the target voice has traditionally been the recognition of normal speech as spoken by average groups of a population. The human factor voice UI defined in this Recommendation can be described as a voice service UI that considers human factors and can handle sound pronunciation difficulties that people may have. When deploying the voice service UI, voice recognition models, post-processing methods and the response UI should be considered depending on the user of the voice UI. Table 1 and clauses 7.2.1 to 7.2.5 give a description of the types of user of the voice UI.

Table 1 – Types of users of the voice UI

Target	Description
The public	A subject with normal pronunciation
Persons with disabilities	Subjects who have problems in pronouncing for inherent or acquired reasons
Patients	Those who suffer from short-term or permanent speech problems due to illness
The elderly	Subjects who have incorrect pronunciation because of old age
Infants & toddlers	Subjects who have not completed the language development process

7.2.1 The public

The general public has been the subject of many speech recognition studies based on normal pronunciation. The user interface for voice services may be developed in terms of convenience, according to the user's state (location, time, event, etc.) according to the application aspect of the user.

7.2.2 The disabled

From the users who have inherited or acquired disabilities, those who can use the voice recognition service should configure the service UI according to the characteristics of the disability.

7.2.3 The patient

The speech UI services used in hospitals or voice UI services for patients with specific diseases should be configured with service UIs considering the service methods and language models that consider the specific disease.

7.2.4 The elderly

With age, pronunciation becomes inaccurate due to problems such as low tone pronunciation, musculoskeletal disorders and dental disorders. Other conditions may also develop with age, for example cognitive disorders, Parkinson's disease, cerebrovascular disease and Alzheimer's disease.

7.2.5 Infants and toddlers

Infants and toddlers are more likely to pronounce more inaccurately than adults, and often communicate in sentences or noun lists with grammatical errors. In this case, it is often difficult to process the voice using a conventional language processing model such those that use grammar-based post-processing.

7.3 Classification according to purpose of use

The voice service can be classified as follows depending on the purpose of use. Various UI types are described in Table 2.

Table 2 – Description of various UI types

UI type	Description
Conversation UI	Voice service UI such as general chatbot
Diagnostic UI	Voice service UI for diagnosing diseases through voice biomarkers
Health UI	Voice service UI used to improve health

7.4 Functional entities

The following is the framework for healthcare voice UI as a key component of e-service. The framework consists of an application layer and a core entities layer, and includes dialogue UI entities for automatic speech recognition (ASR) that connect to the application layer used for the voice UI application. The framework diagram is shown in Figure 4.

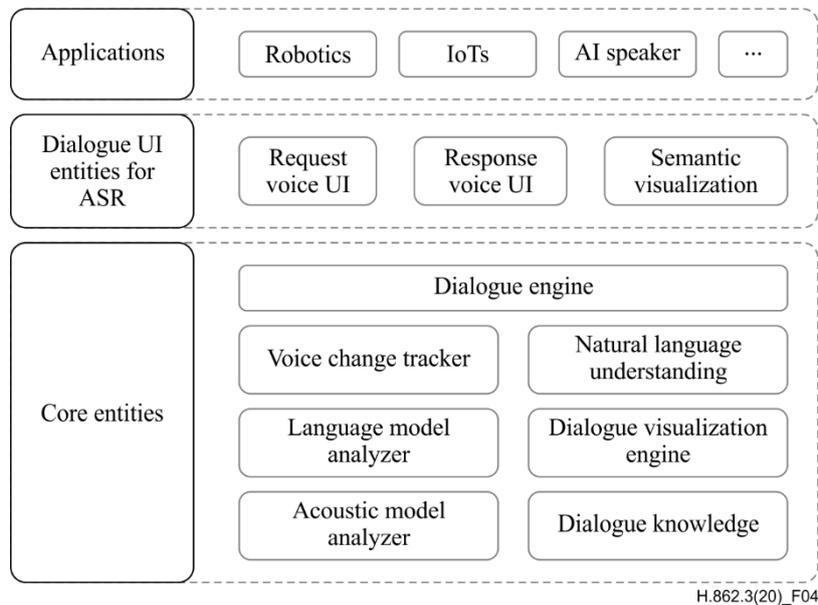


Figure 4 – Voice UI application framework

7.4.1 Dialogue UI

As a general voice user interface for voice conversation, the service UI should be considered for the various users defined in clause 7.2 as well as for general users.

7.4.2 Response UI

The response UI should consider users that can speak using their voice, but whose hearing is impaired, or to have a UI that uses a display or another device for situations where the answer cannot be conveyed by voice. A visual representation of the response in situations where one cannot read or have difficulty in reading can be an alternative means.

7.4.3 Semantic visualization

When there are difficulties hearing voices due to environmental conditions or physical disabilities, the voice UI may provide a method for visualizing and expressing the meaning when the text cannot be typed or cannot be read.

7.4.4 Detection of health change (language model and acoustic model analyser)

Voice can be used as a biomarker to indicate a change in health status. Tracking the ongoing use of an individual's voice service can help identify changes in voice and changes in vocabulary, and can detect changes that help predict conditions such as cardiovascular disease and brain disease.

7.4.5 Knowledge-based conversation for health improvement

Voice services, including chatbots, can improve the health in case of certain diseases such as mental illness.

8 Requirements for the services

8.1 Considerations for service users

The elderly, infants and the disabled have problems such as inaccurate pronunciation because of physical or mental reasons or incorrect grammar usage. As a result, the word error rate (WER) of the sentence is inevitably high. Therefore, the acoustic model (AM) and the language model (LM) normally used for the general public need to be reconstructed taking into consideration pronunciation inaccuracy. Consideration should also be given to the case of a person who can speak but has problems of hearing.

8.1.1 Trigger

A typical voice UI is triggered by users signalling through voice, buttons, gestures, etc. that they want to use speech recognition. In this Recommendation, the user also includes those who cannot use such a voice trigger. Therefore, not only the passive type of trigger but also the way in which the device actively grasps the intention of the user is required to be considered.

8.1.2 Feedback

It is not only necessary to recognize voice and find the right answer, but also to train users based on voice or to predict user behaviour. In addition, the UI is required to be considered taking into consideration the physical circumstances of the user such as the elderly and the disabled.

8.2 Consideration for service environments

The service environment can be a place different from the home, including kindergarten, school, nursing home and hospital. Therefore, the service is required to be robust against noise. The voice-based user recognition may be needed to distinguish the user's voice only.

8.3 Personalized services

The healthcare voice UI should provide speech recognition and conversation services that match the characteristics of the user. To this end, a need to recognize the user and store a history of the user's state is to be considered. In addition, it is considered to be serviced in conjunction with the stored PHRs of the individual.

9 Applications

9.1 Visually impaired user services

Voice UIs can facilitate the use by using the Internet and smart devices. When reading contents from the Internet or a smart phone for users such as the visually impaired, or those affected by factors such as amblyopia and dyslexia, the voice UI can convert the text into voice and speak it out to the user. In addition, it is possible to provide a function to recognize and provide a vocal description of the object. A block diagram of the visually impaired user services is shown in Figure 5.

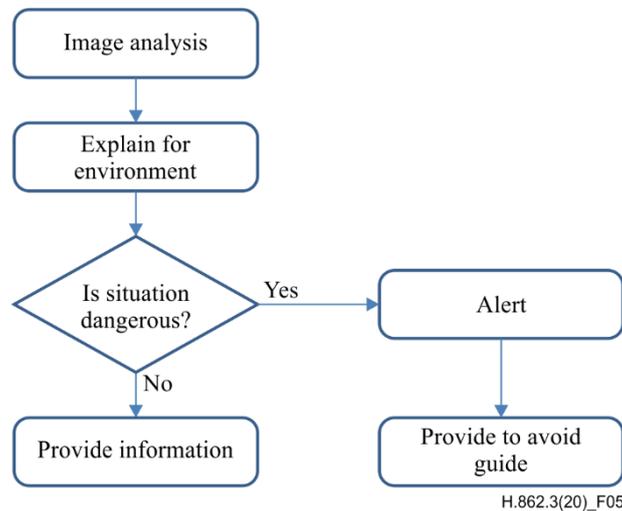


Figure 5 – Block diagram of visually impaired user services

9.2 The voice UI application for healthcare services

The voice UI service is capable of continuously tracking the voices of the elderly and connecting them to the necessary healthcare services. Tracking and diagnosing the persistent voice changes presented in this Recommendation will allow for prediction of emergencies and early disease, enabling users to detect health problems and provide a quick response. Figure 6 shows an example of the voice UI application for elderly healthcare services.

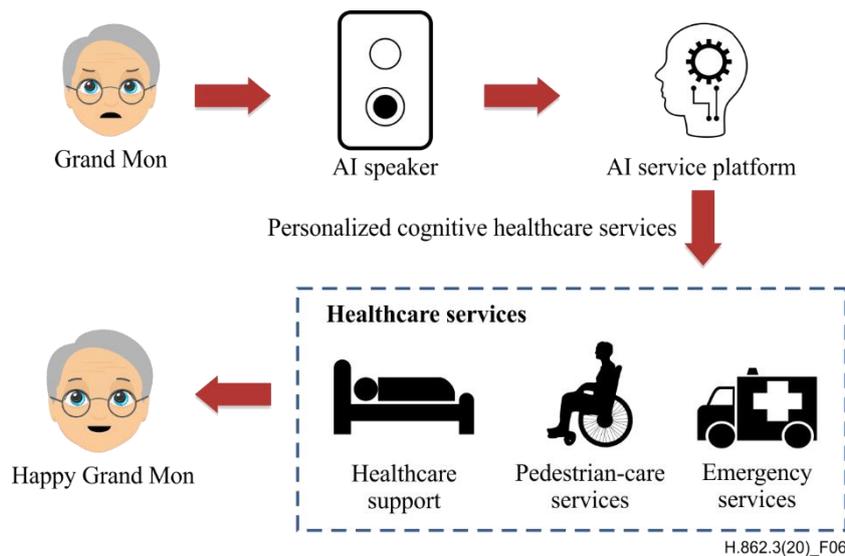


Figure 6 – Example of the voice UI application for elderly healthcare services

9.3 Health improvement and promotion for elderly

The increment of life expectancy is increasing the number of elderly people living alone. Seniors living alone may have emotional or psychological problems, such as depression. It is possible to solve these various problems of the elderly through a voice user interface. The following are examples of voice UI services:

- Psychological care through dialogue
- Dosage management using voice UI of elderly patients with chronic diseases
- Check health status through voice UI

Bibliography

- [b-ITU-T F.746.3] Recommendation ITU-T F.746.3 (2015), *Intelligent question answering service framework*.
- [b-ITU-T H.703] Recommendation ITU-T H.703 (2016), *Enhanced user interface framework for IPTV terminal devices*.

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