US 20250259633A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2025/0259633 A1**

Lee et al. (43) **Pub. Date: Aug. 14, 2025**

(54) **ELECTRONIC DEVICE, SERVER, AND SYSTEM INCLUDING SAME**

(71) Applicant: **LG ELECTRONICS INC.**, Seoul (KR)

(72) Inventors: **Seonghak Lee**, Seoul (KR); **Gowoon Choi**, Seoul (KR); **Sangseok Lee**, Seoul (KR); **Taejoon Kim**, Seoul (KR); **Seunghye Park**, Seoul (KR); **Jeongyong Park**, Seoul (KR); **Jungjae Lee**, Seoul (KR)

(73) Assignee: **LG ELECTRONICS INC.**, Seoul (KR)

(21) Appl. No.: **19/048,010**

(22) Filed: **Feb. 7, 2025**

(30) **Foreign Application Priority Data**

Feb. 8, 2024 (KR) 10-2024-0019974
Jul. 31, 2024 (WO) PCT/KR2024/011228

Publication Classification

(51) **Int. Cl.**
G10L 17/22 (2013.01)
G10L 15/26 (2006.01)
G10L 17/02 (2013.01)

(52) **U.S. Cl.**
CPC ***G10L 17/22*** (2013.01); ***G10L 15/26*** (2013.01); ***G10L 17/02*** (2013.01)

(57) **ABSTRACT**

An electronic device includes a display, an external device interface that communicates with a remote control device, a network interface that communicates with a server, a user input interface that transmits a signal related to a user input, and a controller, and the controller displays a preset text related to registration of identification information while performing a process of registering the identification information related to speech for a user account logged in to the server, upon receiving a speech signal related to the preset text, transmits data including the speech signal to the server, completes the process of registering the identification information based on processing of the speech signal related to the preset text by the server, and interrupts the process of registering the identification information upon receiving a predetermined input related to speech recognition from the remote control device during the process of registering the identification information.

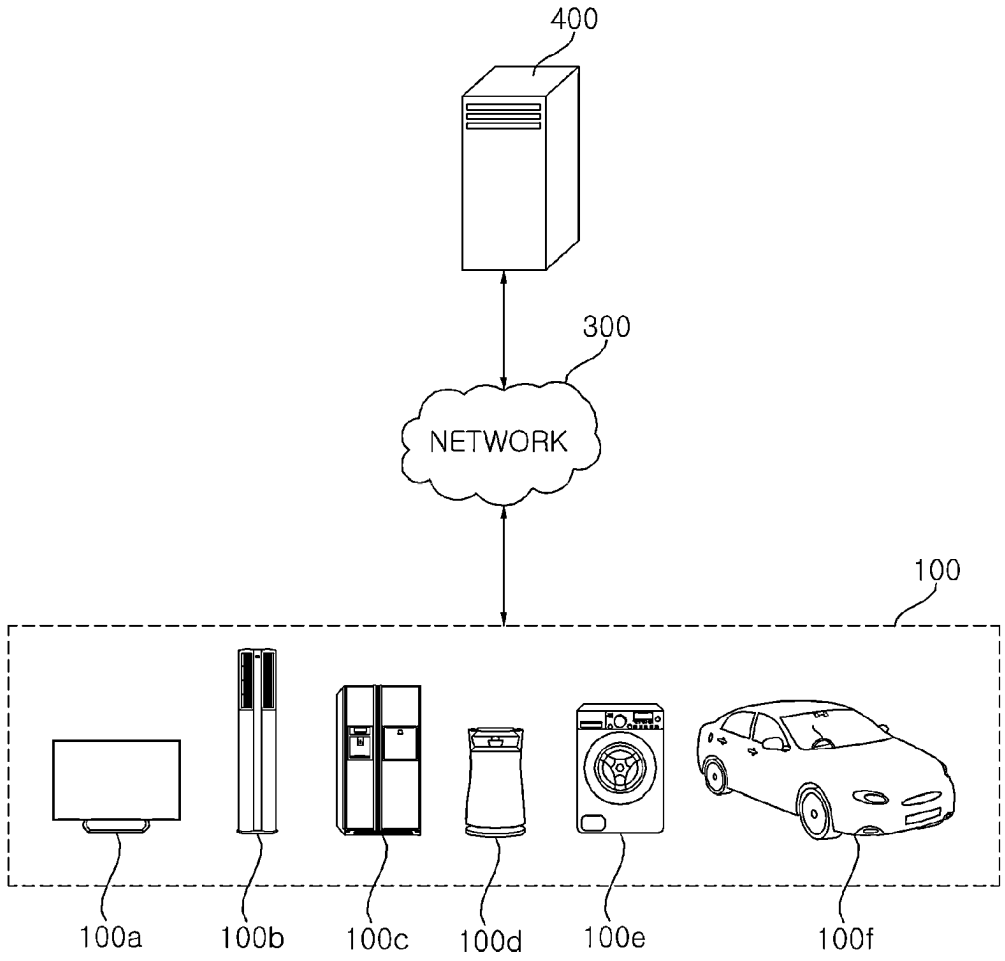


FIG. 1

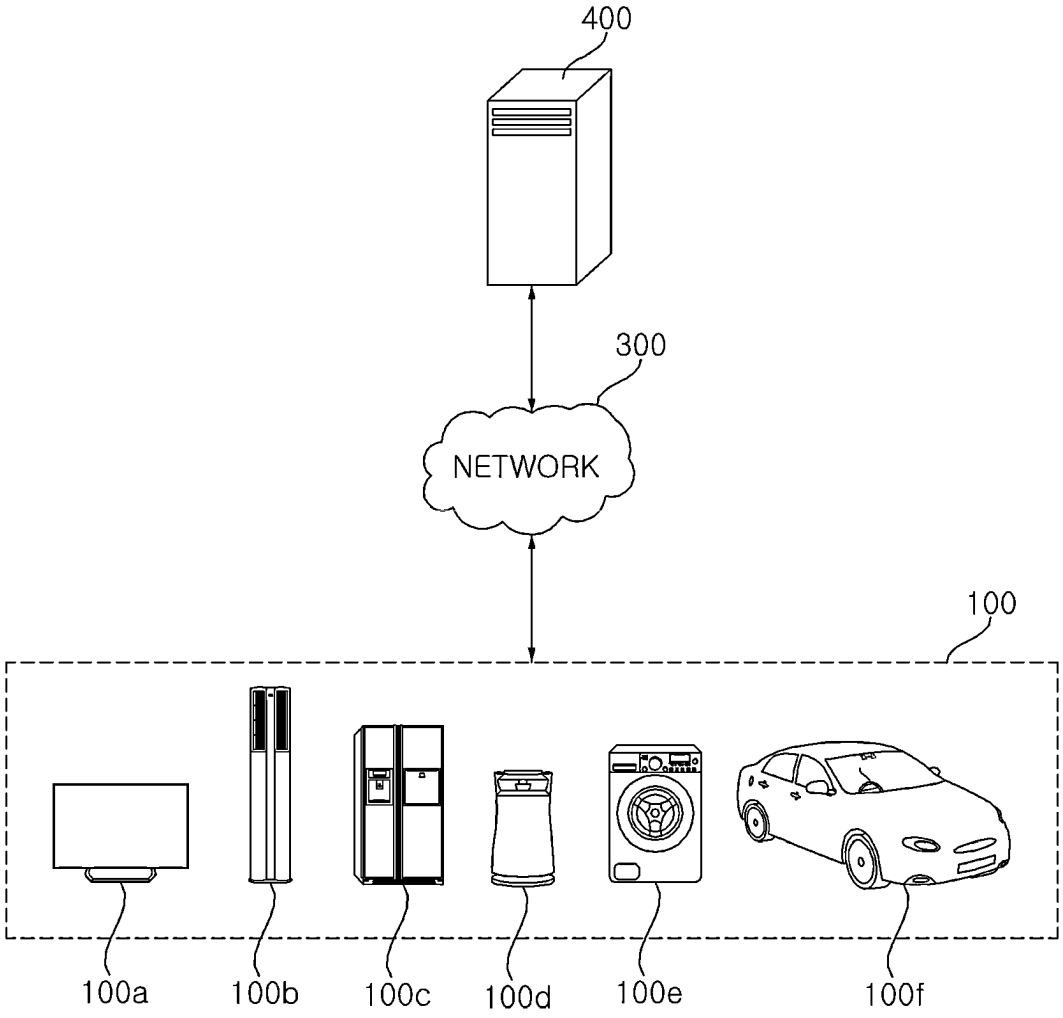


FIG. 2

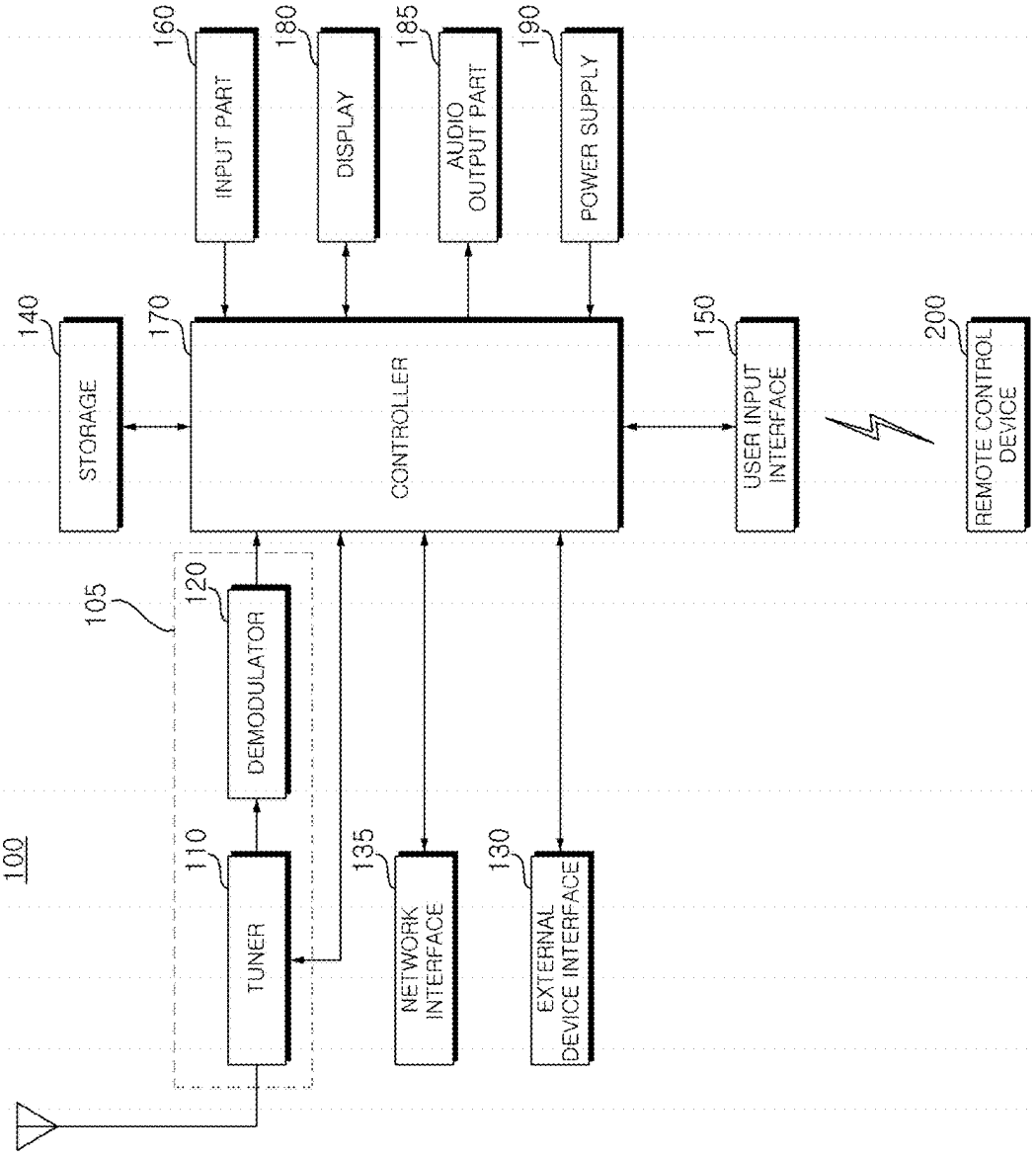

100
105
110
TUNER
120
DEMODULATOR
130
EXTERNAL DEVICE INTERFACE
135
NETWORK INTERFACE
140
STORAGE
150
USER INPUT INTERFACE
160
INPUT PART
170
CONTROLLER
180
DISPLAY
185
AUDIO OUTPUT PART
190
POWER SUPPLY
200
REMOTE CONTROL DEVICE

FIG. 3

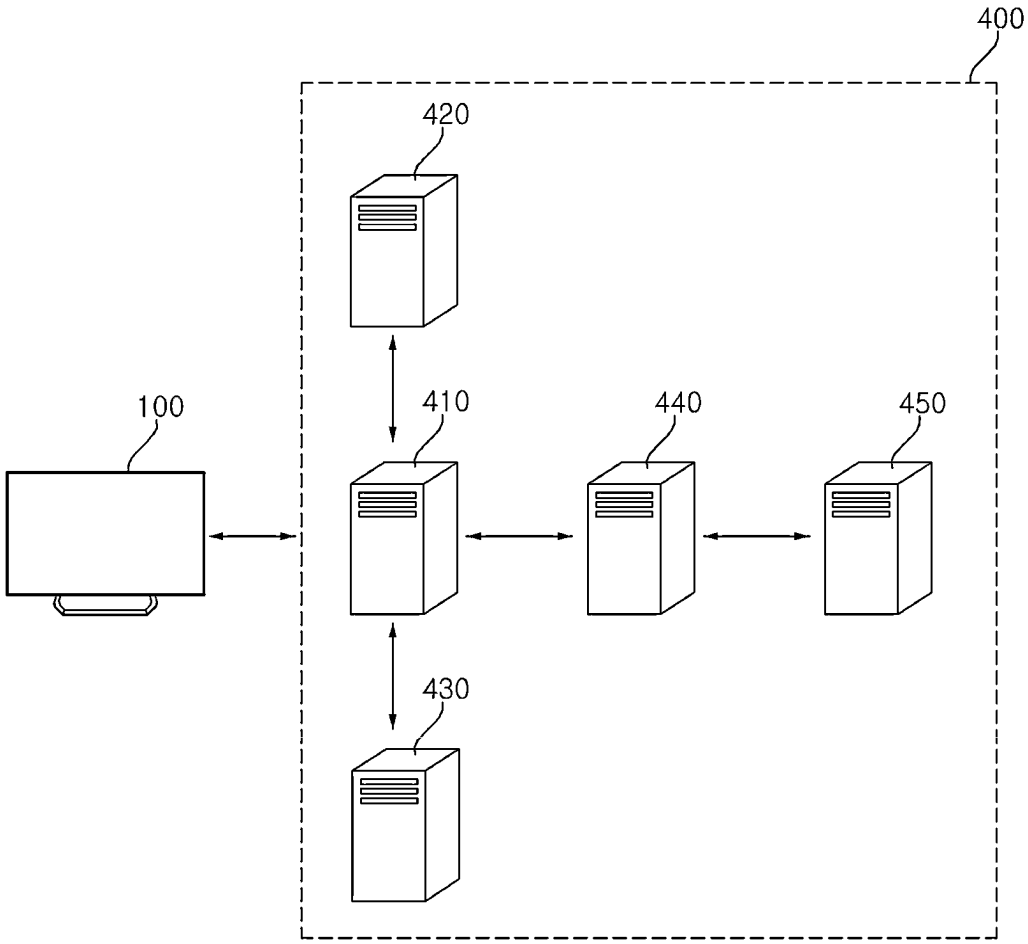


FIG. 4

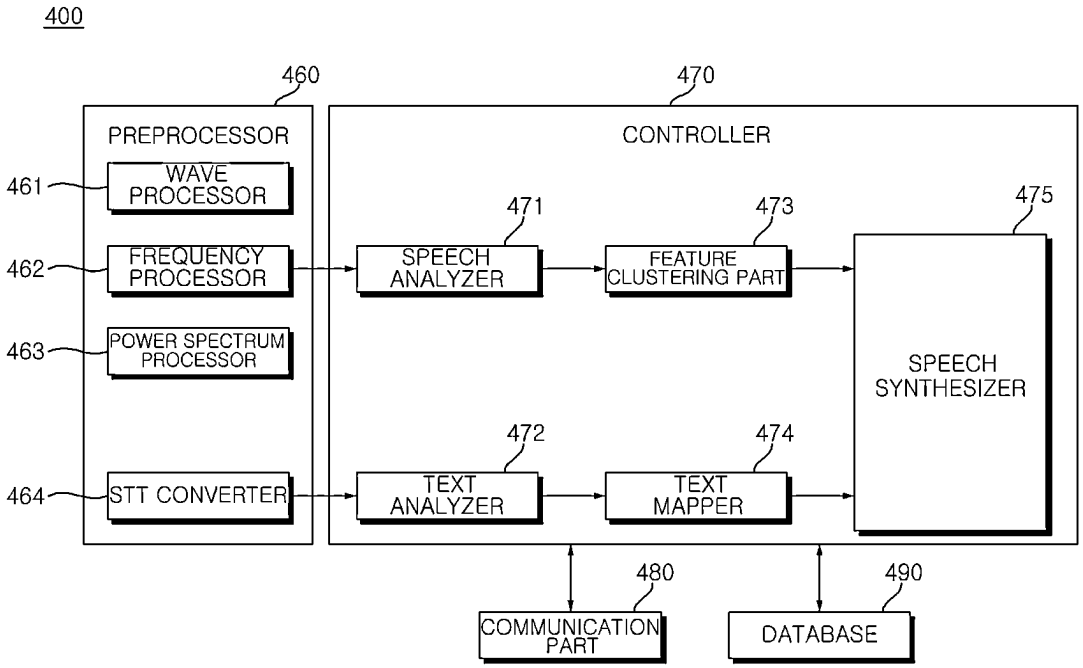

400
460
470
PREPROCESSOR
CONTROLLER
461
WAVE PROCESSOR
462
FREQUENCY PROCESSOR
463
POWER SPECTRUM PROCESSOR
464
STT CONVERTER
471
SPEECH ANALYZER
473
FEATURE CLUSTERING PART
475
SPEECH SYNTHESIZER
472
TEXT ANALYZER
474
TEXT MAPPER
480
COMMUNICATION PART
490
DATABASE

FIG. 5

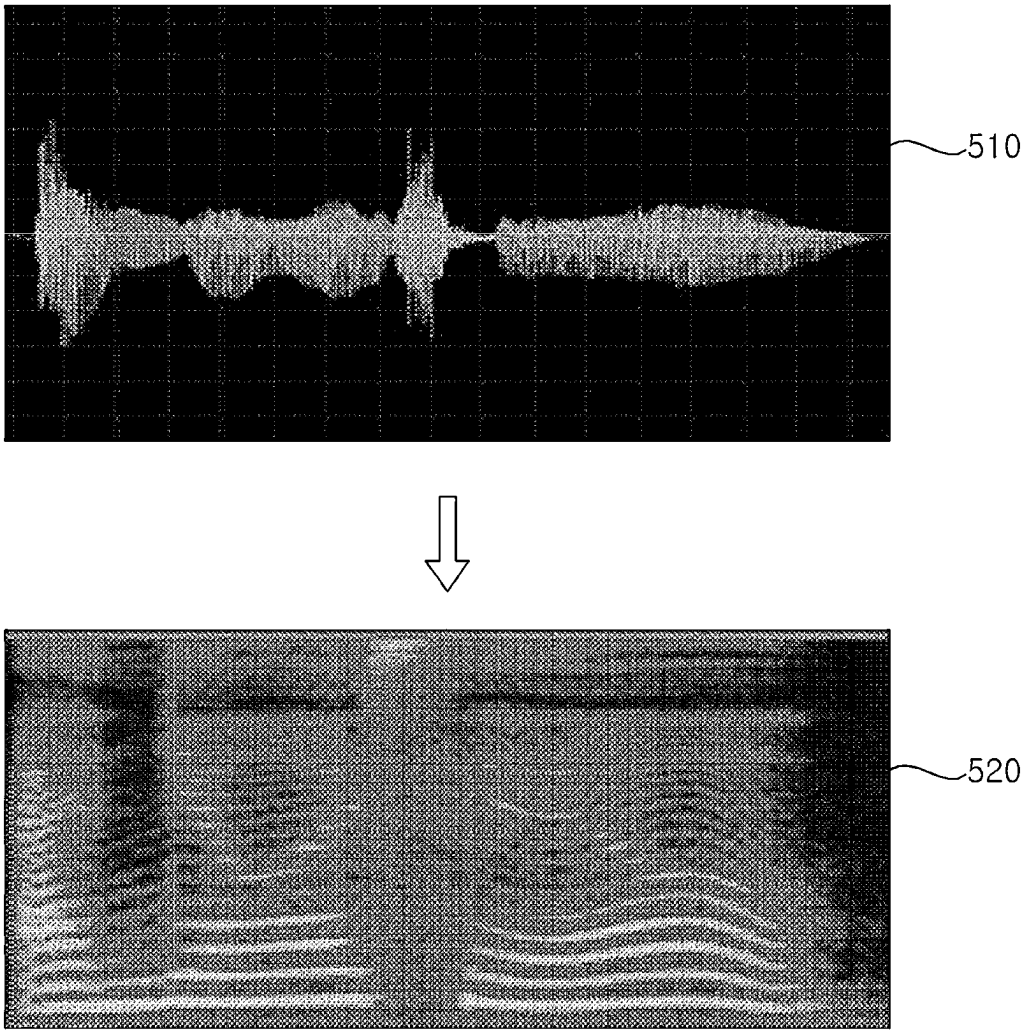


FIG. 6

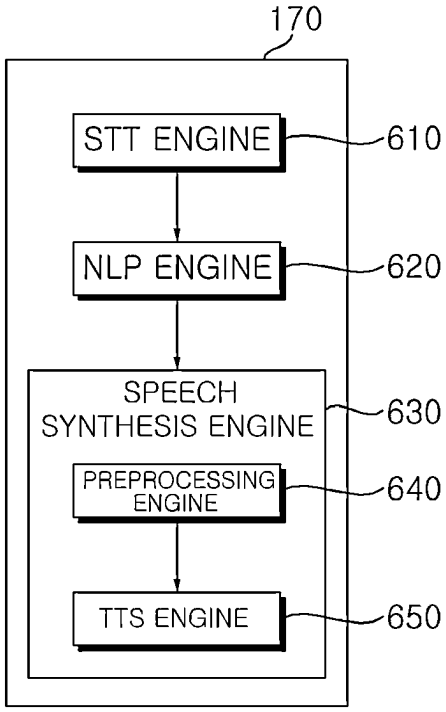

170
STT ENGINE
610
NLP ENGINE
620
SPEECH SYNTHESIS ENGINE
630
PREPROCESSING ENGINE
640
TTS ENGINE
650

FIG. 7

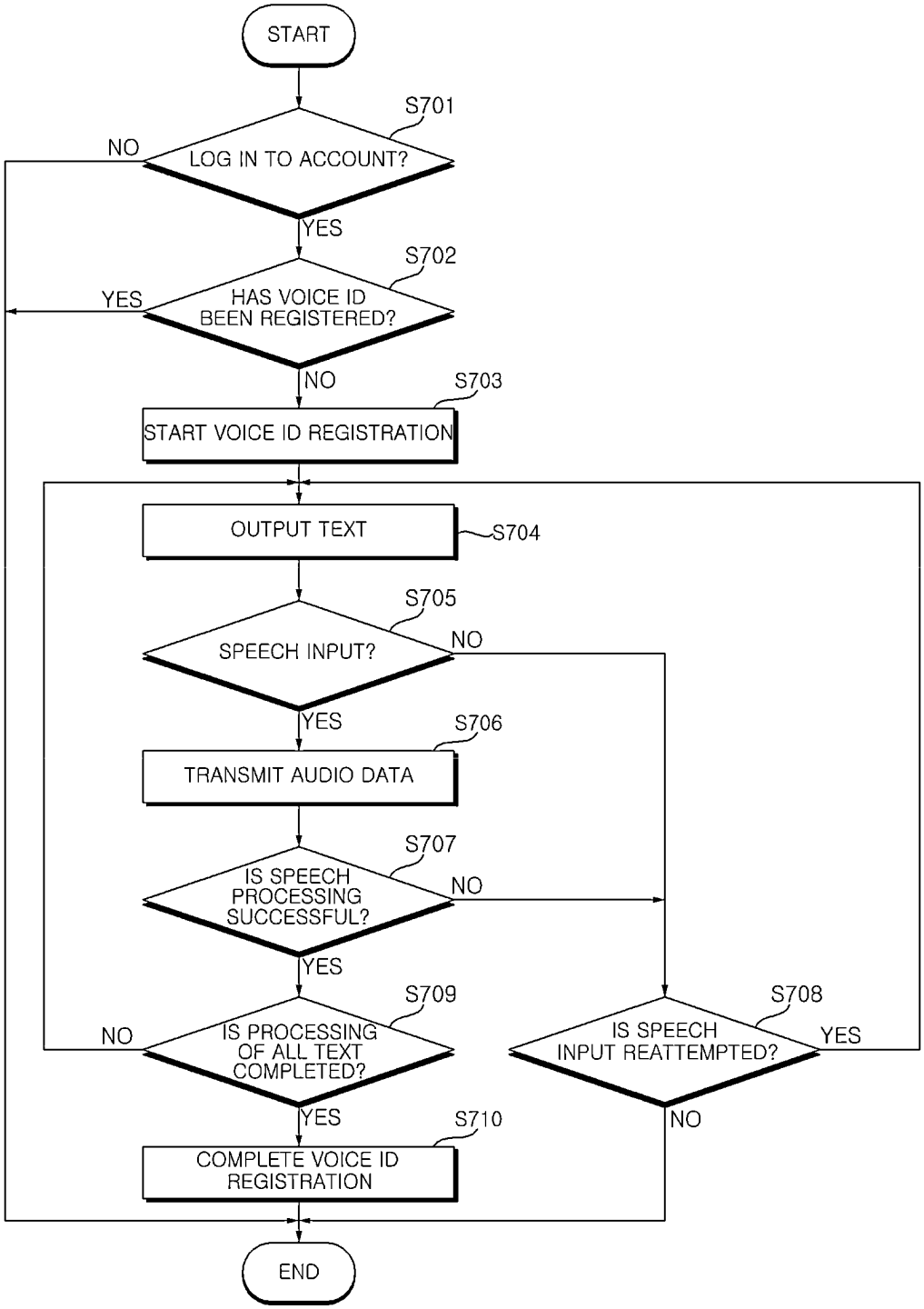

START
S701
LOG IN TO ACCOUNT?
NO
YES
S702
HAS VOICE ID BEEN REGISTERED?
YES
NO
S703
START VOICE ID REGISTRATION
OUTPUT TEXT
S704
S705
SPEECH INPUT?
NO
YES
S706
TRANSMIT AUDIO DATA
S707
IS SPEECH PROCESSING SUCCESSFUL?
NO
YES
S709
IS PROCESSING OF ALL TEXT COMPLETED?
NO
YES
S708
IS SPEECH INPUT REATTEMPTED?
YES
NO
S710
COMPLETE VOICE ID REGISTRATION
END

FIG. 8

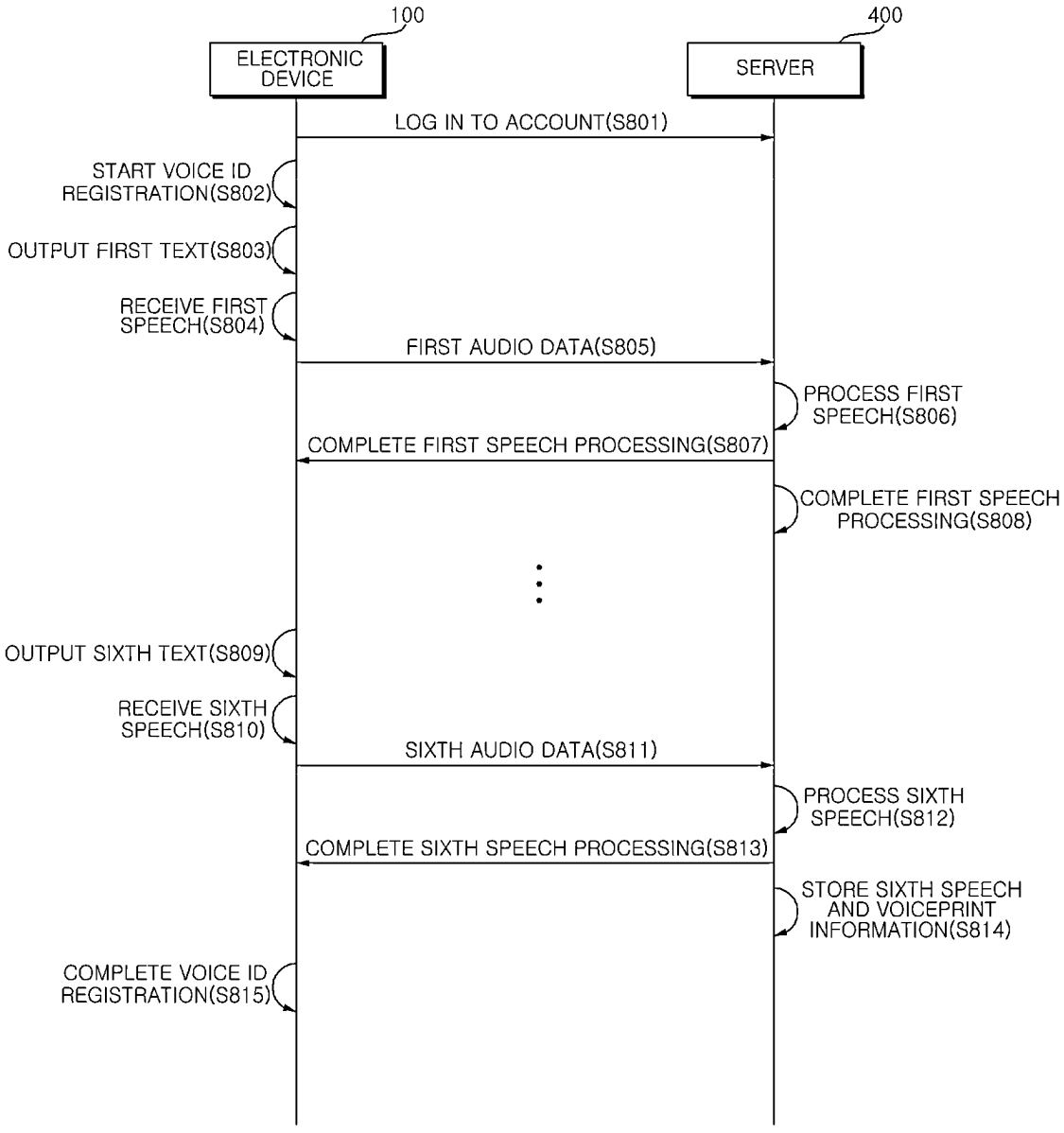

100
ELECTRONIC DEVICE
400
SERVER
LOG IN TO ACCOUNT(S801)
START VOICE ID REGISTRATION(S802)
OUTPUT FIRST TEXT(S803)
RECEIVE FIRST SPEECH(S804)
FIRST AUDIO DATA(S805)
PROCESS FIRST SPEECH(S806)
COMPLETE FIRST SPEECH PROCESSING(S807)
COMPLETE FIRST SPEECH PROCESSING(S808)
OUTPUT SIXTH TEXT(S809)
RECEIVE SIXTH SPEECH(S810)
SIXTH AUDIO DATA(S811)
PROCESS SIXTH SPEECH(S812)
COMPLETE SIXTH SPEECH PROCESSING(S813)
STORE SIXTH SPEECH AND VOICEPRINT INFORMATION(S814)
COMPLETE VOICE ID REGISTRATION(S815)

FIG. 9

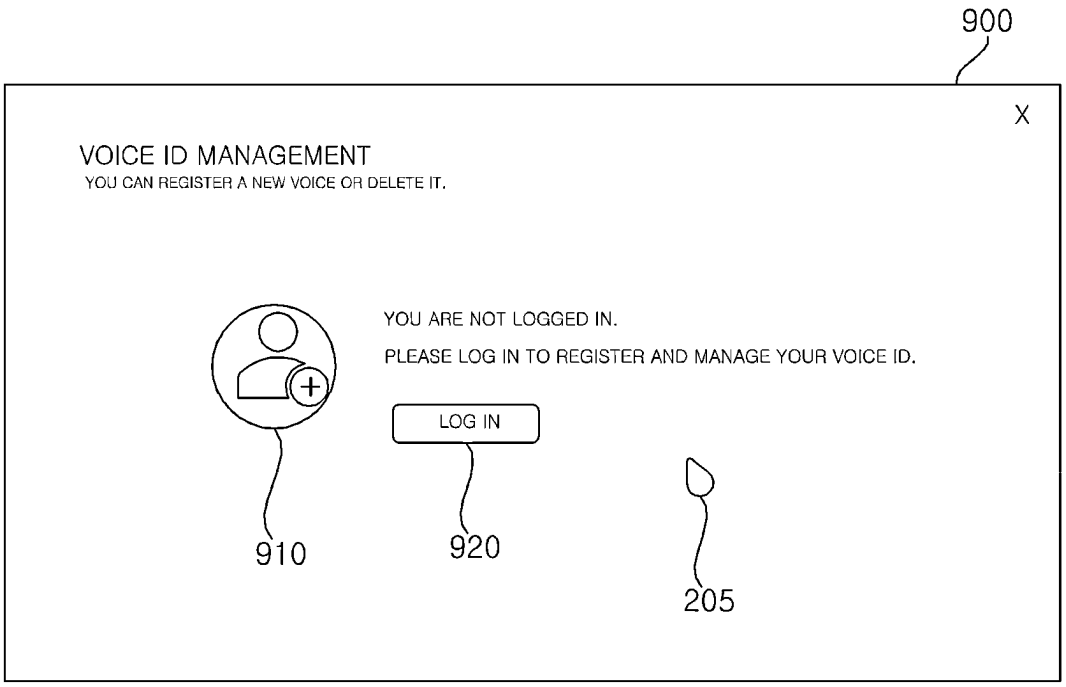

900
X
VOICE ID MANAGEMENT
YOU CAN REGISTER A NEW VOICE OR DELETE IT.
YOU ARE NOT LOGGED IN.
PLEASE LOG IN TO REGISTER AND MANAGE YOUR VOICE ID.
LOG IN
910
920
205

FIG. 10

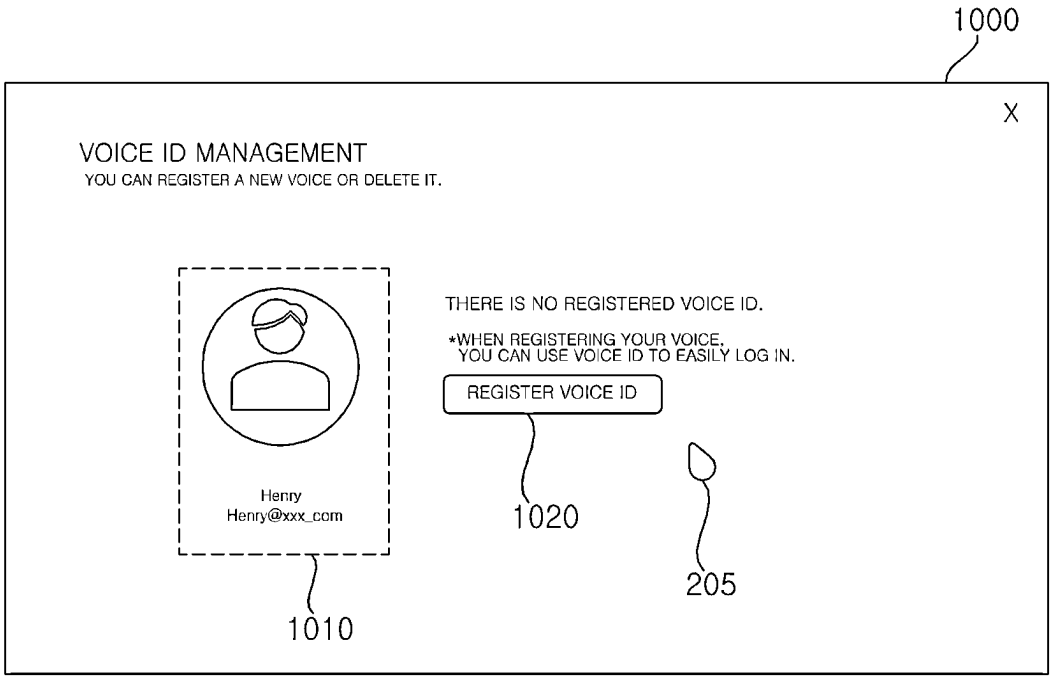

1000
X
VOICE ID MANAGEMENT
YOU CAN REGISTER A NEW VOICE OR DELETE IT.
Henry
Henry@xxx_com
1010
THERE IS NO REGISTERED VOICE ID.
*WHEN REGISTERING YOUR VOICE,
YOU CAN USE VOICE ID TO EASILY LOG IN.
REGISTER VOICE ID
1020
205

FIG. 11A

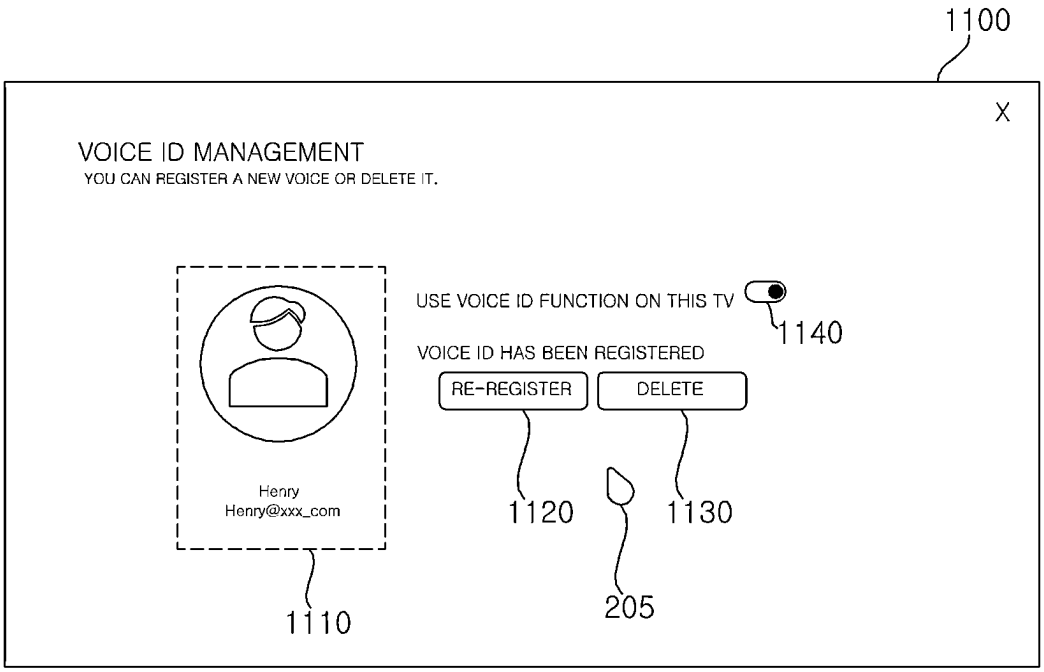

1100
X
VOICE ID MANAGEMENT
YOU CAN REGISTER A NEW VOICE OR DELETE IT.
Henry
Henry@xxx_com
1110
USE VOICE ID FUNCTION ON THIS TV
1140
VOICE ID HAS BEEN REGISTERED
RE-REGISTER
DELETE
1120
1130
205

FIG. 11B

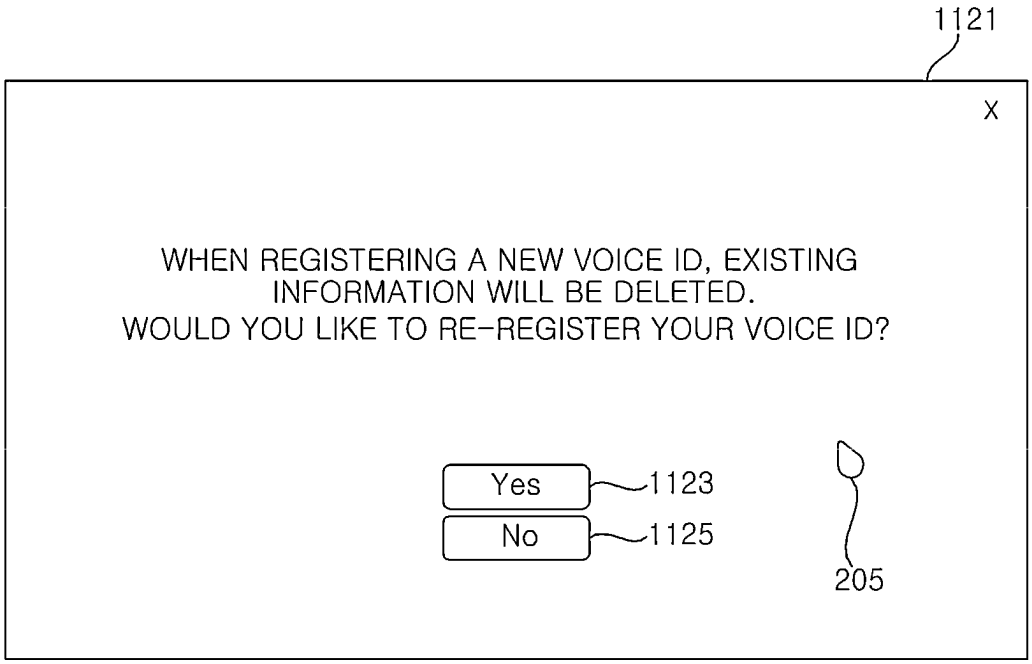

1121
X
WHEN REGISTERING A NEW VOICE ID, EXISTING INFORMATION WILL BE DELETED.
WOULD YOU LIKE TO RE-REGISTER YOUR VOICE ID?
Yes
1123
No
1125
205

FIG. 11C

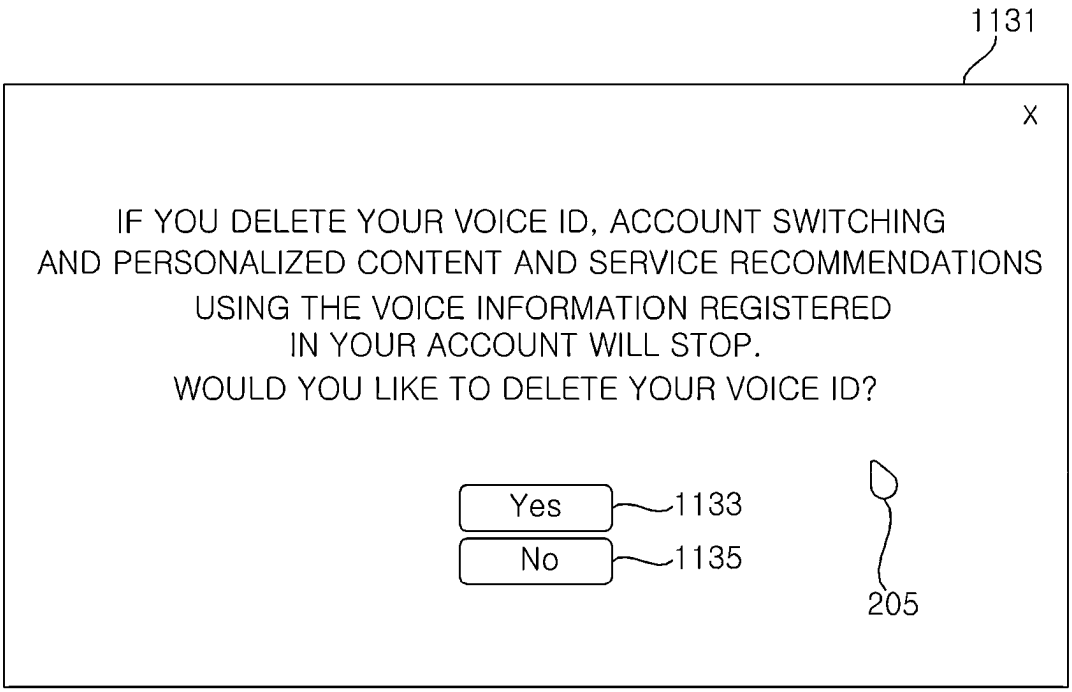

1131
X
IF YOU DELETE YOUR VOICE ID, ACCOUNT SWITCHING
AND PERSONALIZED CONTENT AND SERVICE RECOMMENDATIONS
USING THE VOICE INFORMATION REGISTERED
IN YOUR ACCOUNT WILL STOP.
WOULD YOU LIKE TO DELETE YOUR VOICE ID?
Yes
1133
No
1135
205

FIG. 12

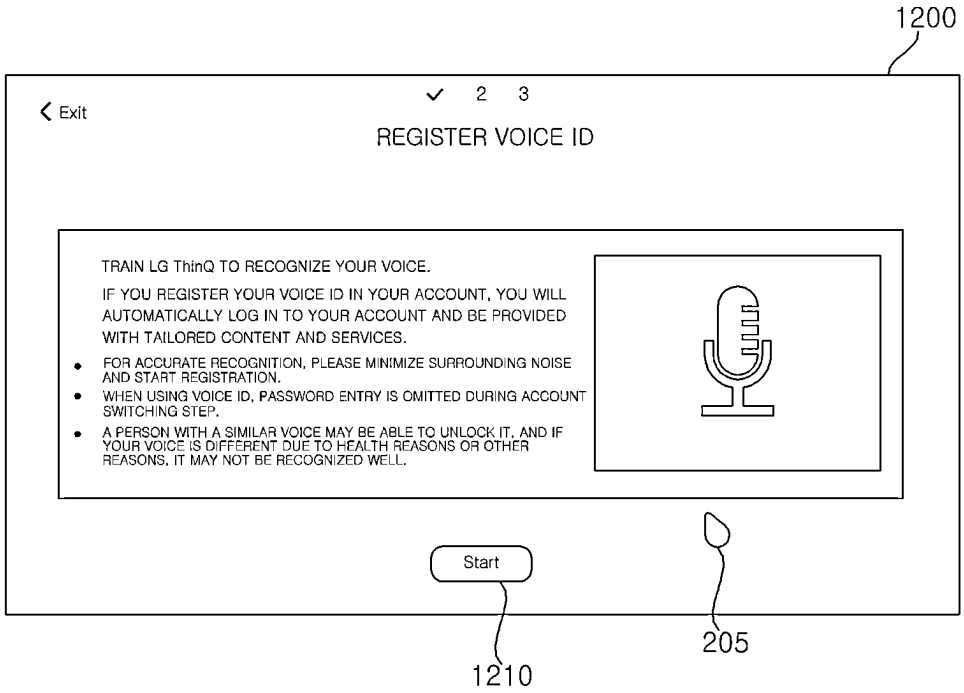

1200
Exit
2 3
REGISTER VOICE ID
TRAIN LG ThinQ TO RECOGNIZE YOUR VOICE.
IF YOU REGISTER YOUR VOICE ID IN YOUR ACCOUNT, YOU WILL AUTOMATICALLY LOG IN TO YOUR ACCOUNT AND BE PROVIDED WITH TAILORED CONTENT AND SERVICES.
FOR ACCURATE RECOGNITION, PLEASE MINIMIZE SURROUNDING NOISE AND START REGISTRATION.
WHEN USING VOICE ID, PASSWORD ENTRY IS OMITTED DURING ACCOUNT SWITCHING STEP.
A PERSON WITH A SIMILAR VOICE MAY BE ABLE TO UNLOCK IT, AND IF YOUR VOICE IS DIFFERENT DUE TO HEALTH REASONS OR OTHER REASONS, IT MAY NOT BE RECOGNIZED WELL.
Start
205
1210

FIG. 13

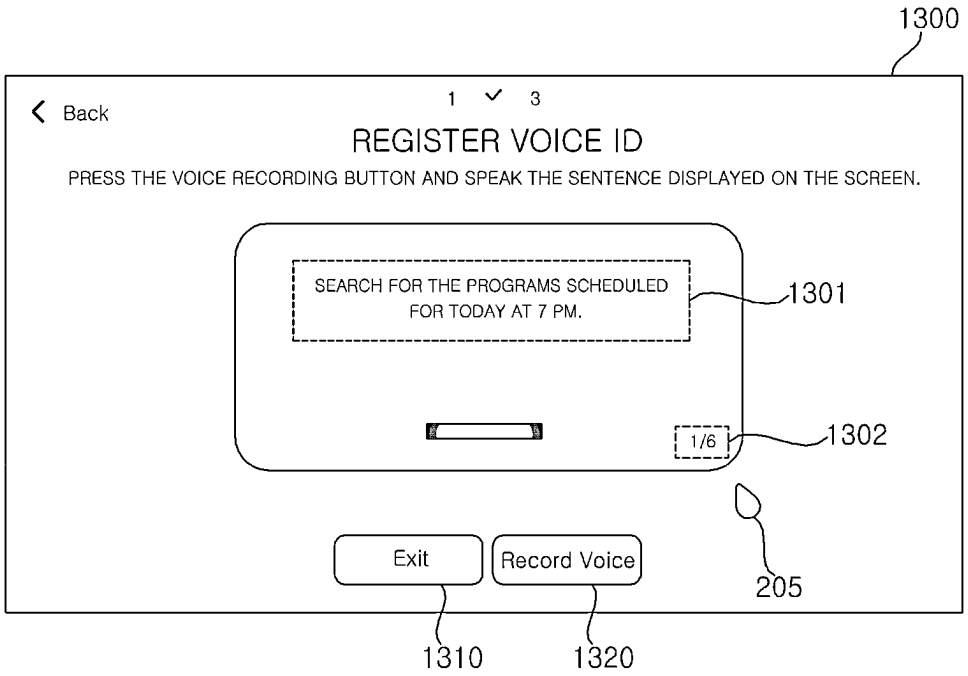

1300
Back
1 3
REGISTER VOICE ID
PRESS THE VOICE RECORDING BUTTON AND SPEAK THE SENTENCE DISPLAYED ON THE SCREEN.
SEARCH FOR THE PROGRAMS SCHEDULED FOR TODAY AT 7 PM.
1301
1/6
1302
Exit
Record Voice
205
1310
1320

FIG. 14

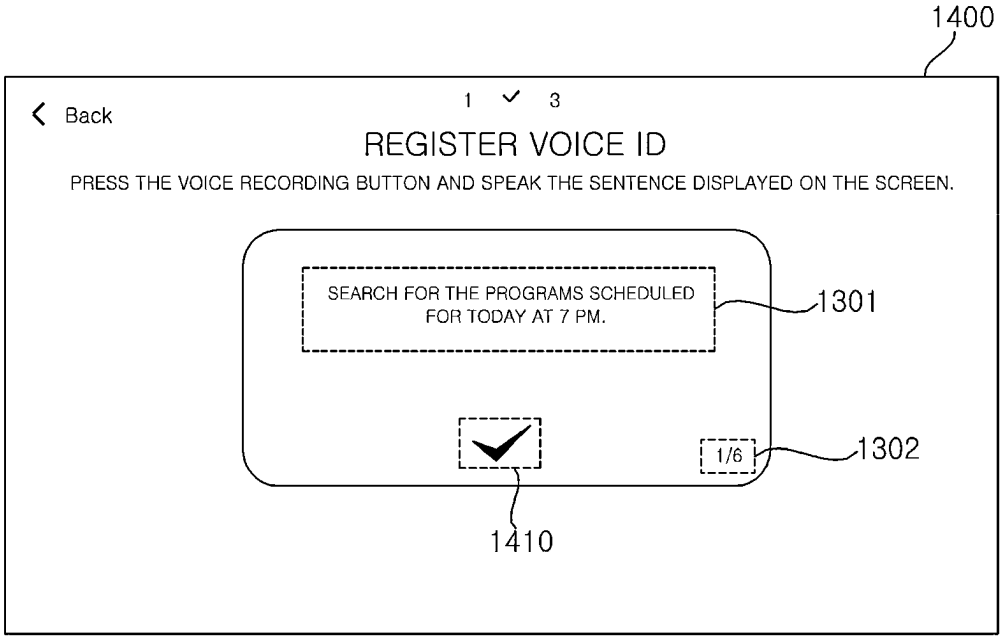

1400
Back
1 3
REGISTER VOICE ID
PRESS THE VOICE RECORDING BUTTON AND SPEAK THE SENTENCE DISPLAYED ON THE SCREEN.
SEARCH FOR THE PROGRAMS SCHEDULED FOR TODAY AT 7 PM.
1301
1/6
1302
1410

FIG. 15

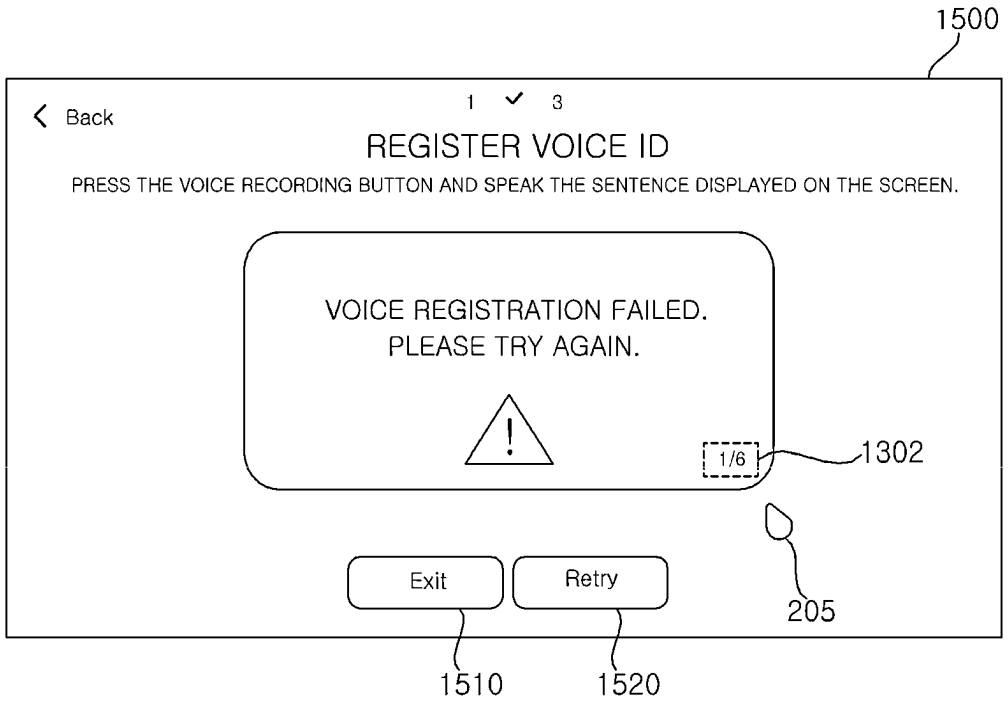

1500
Back
1 3
REGISTER VOICE ID
PRESS THE VOICE RECORDING BUTTON AND SPEAK THE SENTENCE DISPLAYED ON THE SCREEN.
VOICE REGISTRATION FAILED.
PLEASE TRY AGAIN.
1/6
1302
205
Exit
Retry
1510
1520

FIG. 16

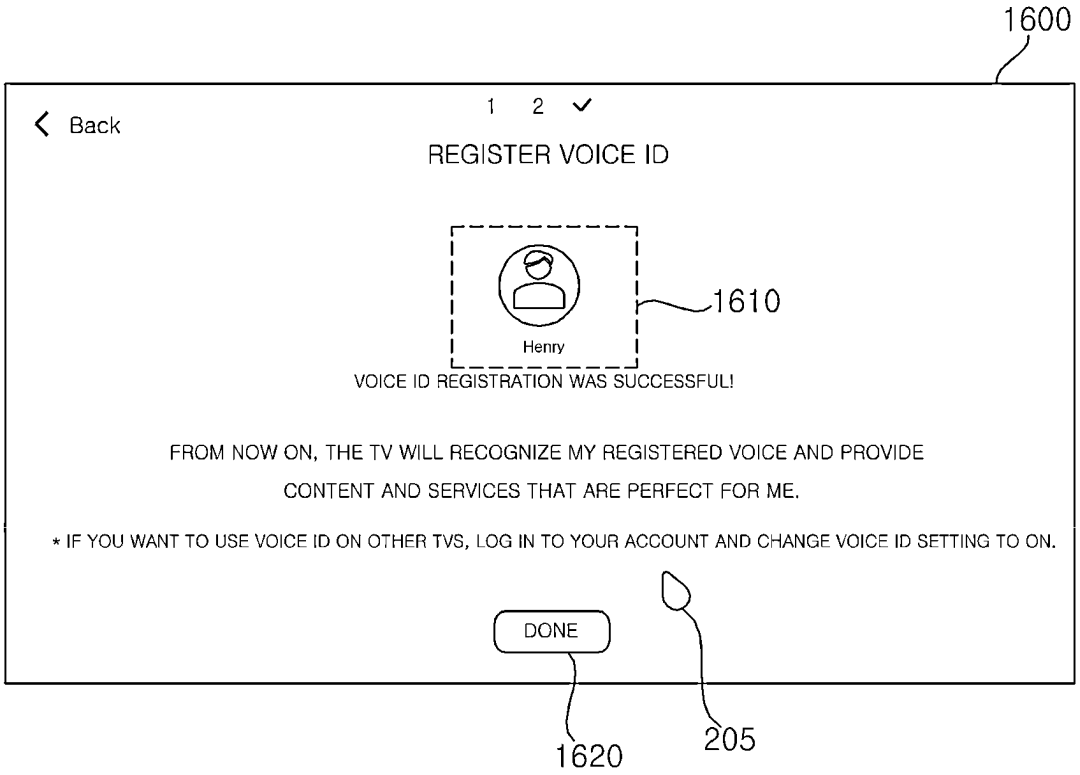

1600
Back
1 2
REGISTER VOICE ID
1610
Henry
VOICE ID REGISTRATION WAS SUCCESSFUL!
FROM NOW ON, THE TV WILL RECOGNIZE MY REGISTERED VOICE AND PROVIDE
CONTENT AND SERVICES THAT ARE PERFECT FOR ME.
* IF YOU WANT TO USE VOICE ID ON OTHER TVS, LOG IN TO YOUR ACCOUNT AND CHANGE VOICE ID SETTING TO ON.
DONE
1620
205

FIG. 17A

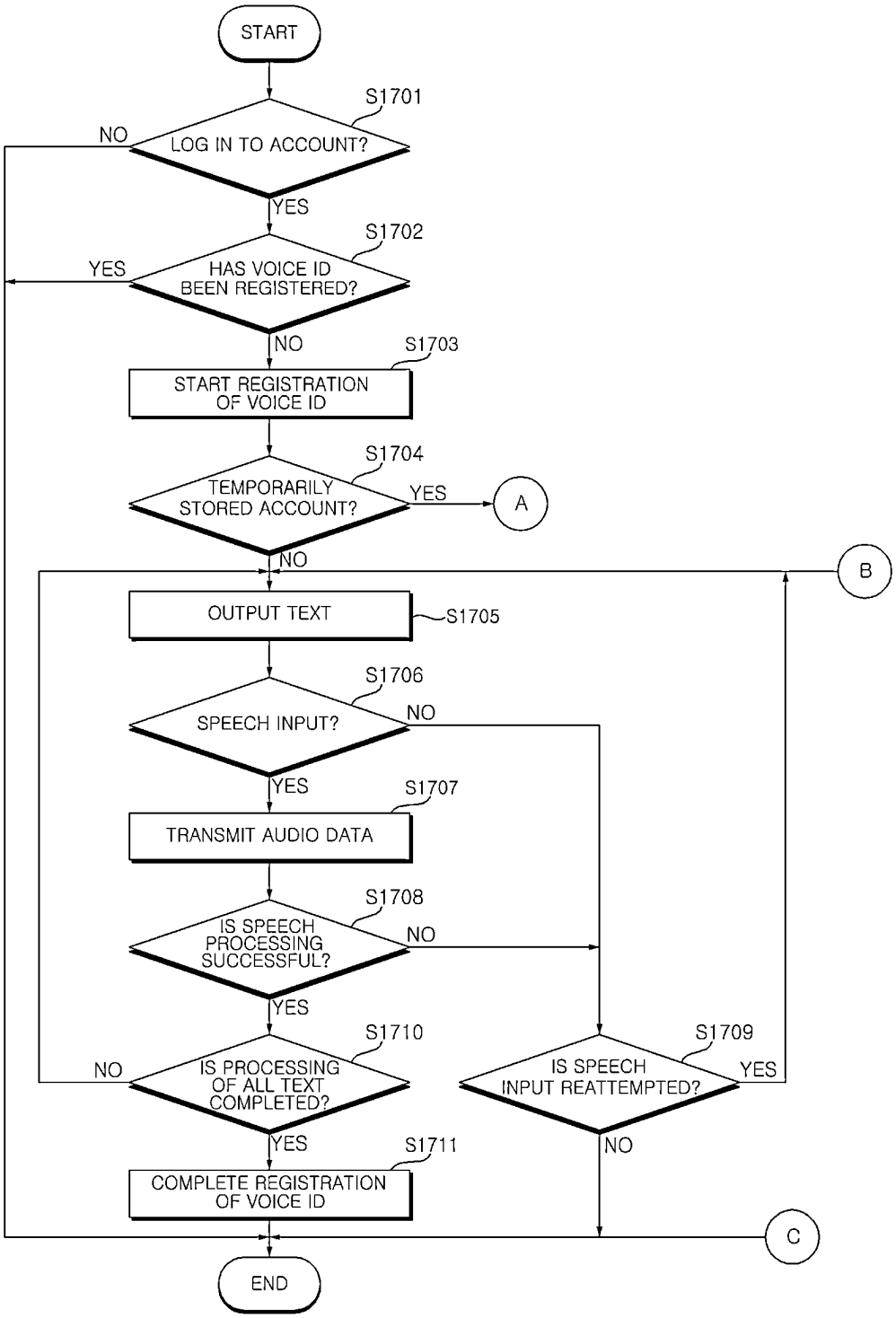

START
S1701
LOG IN TO ACCOUNT?
NO
YES
S1702
HAS VOICE ID BEEN REGISTERED?
YES
NO
S1703
START REGISTRATION OF VOICE ID
S1704
TEMPORARILY STORED ACCOUNT?
YES
A
NO
B
OUTPUT TEXT
S1705
S1706
SPEECH INPUT?
NO
YES
S1707
TRANSMIT AUDIO DATA
S1708
IS SPEECH PROCESSING SUCCESSFUL?
NO
YES
S1710
IS PROCESSING OF ALL TEXT COMPLETED?
NO
S1709
IS SPEECH INPUT REATTEMPTED?
YES
YES
NO
S1711
COMPLETE REGISTRATION OF VOICE ID
C
END

FIG. 17B

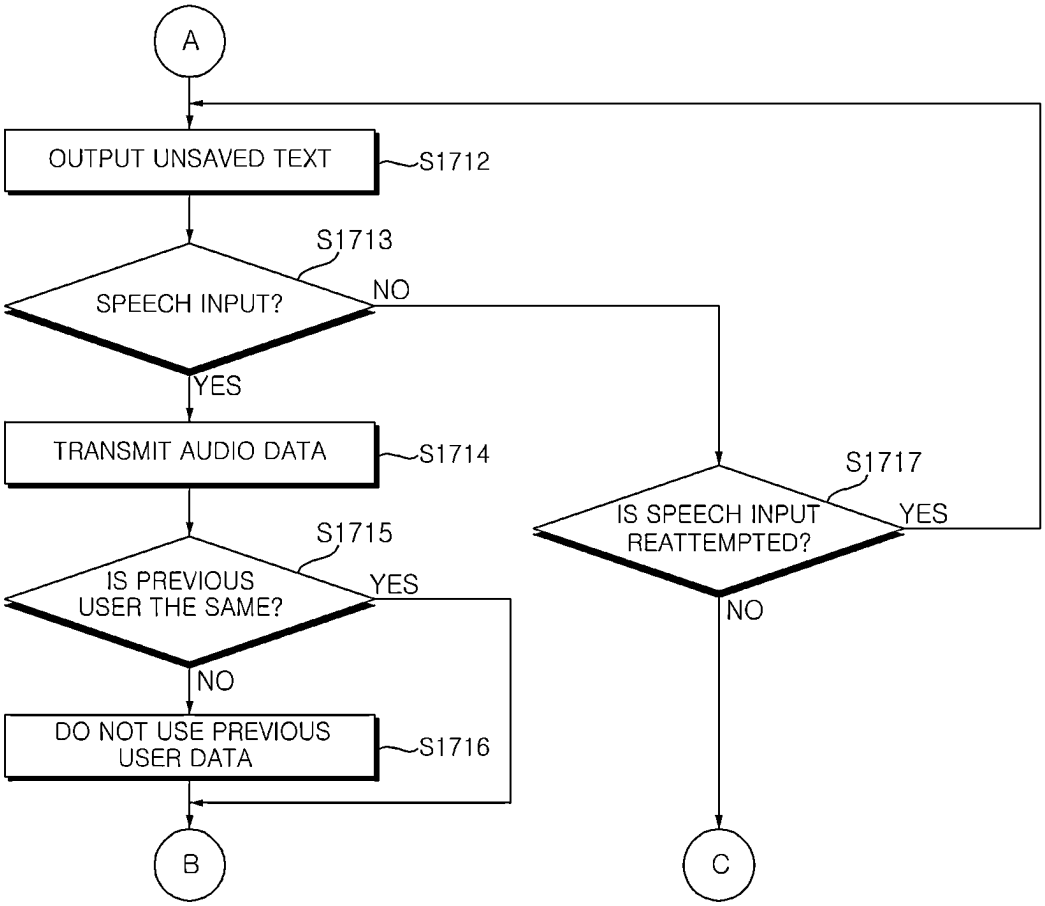

A
OUTPUT UNSAVED TEXT
S1712
S1713
SPEECH INPUT?
NO
YES
TRANSMIT AUDIO DATA
S1714
S1715
IS PREVIOUS USER THE SAME?
YES
NO
DO NOT USE PREVIOUS USER DATA
S1716
B
S1717
IS SPEECH INPUT REATTEMPTED?
YES
NO
C

FIG. 18

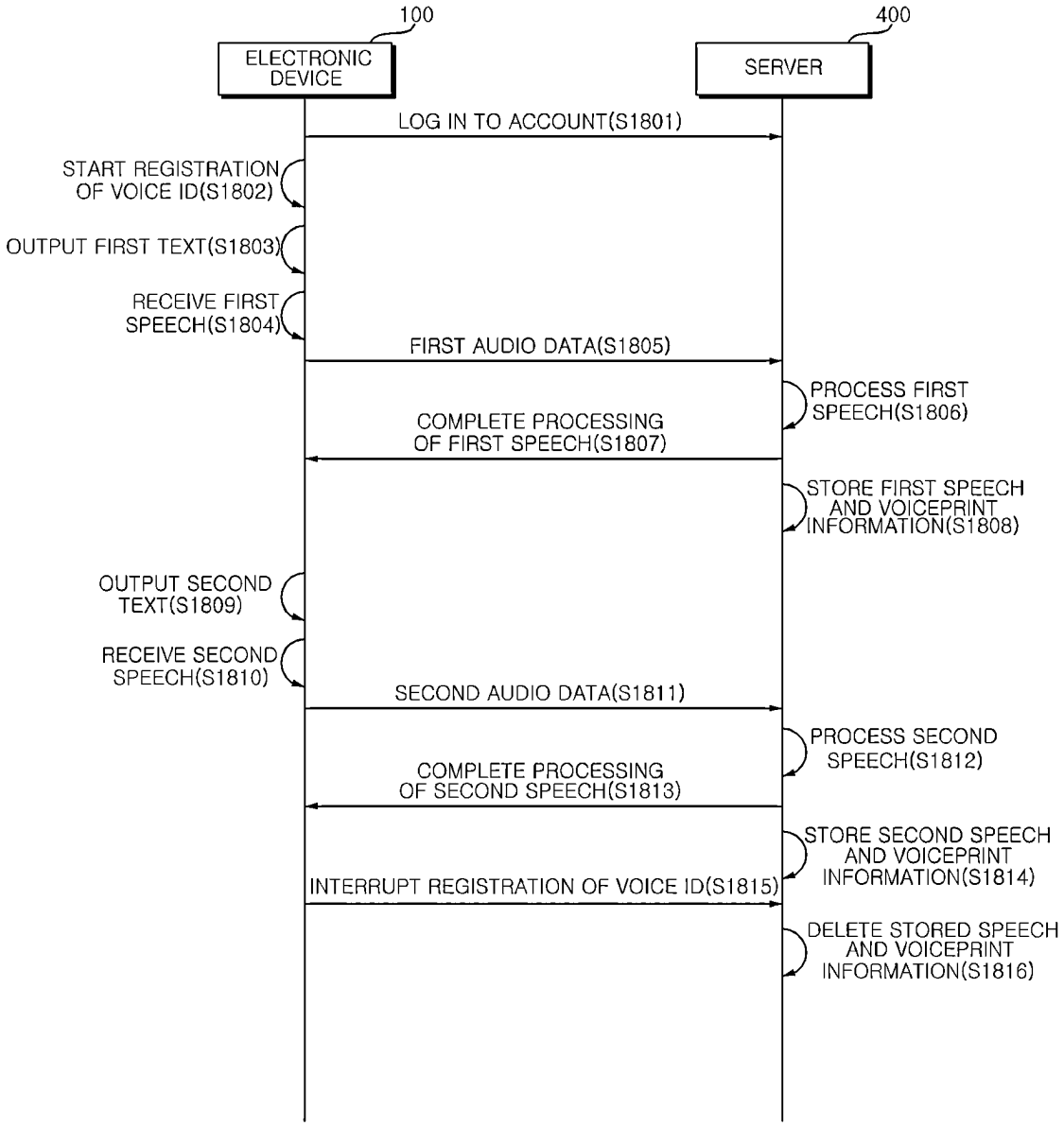

100
ELECTRONIC DEVICE
400
SERVER
LOG IN TO ACCOUNT(S1801)
START REGISTRATION OF VOICE ID(S1802)
OUTPUT FIRST TEXT(S1803)
RECEIVE FIRST SPEECH(S1804)
FIRST AUDIO DATA(S1805)
PROCESS FIRST SPEECH(S1806)
COMPLETE PROCESSING OF FIRST SPEECH(S1807)
STORE FIRST SPEECH AND VOICEPRINT INFORMATION(S1808)
OUTPUT SECOND TEXT(S1809)
RECEIVE SECOND SPEECH(S1810)
SECOND AUDIO DATA(S1811)
PROCESS SECOND SPEECH(S1812)
COMPLETE PROCESSING OF SECOND SPEECH(S1813)
STORE SECOND SPEECH AND VOICEPRINT INFORMATION(S1814)
INTERRUPT REGISTRATION OF VOICE ID(S1815)
DELETE STORED SPEECH AND VOICEPRINT INFORMATION(S1816)

FIG. 19

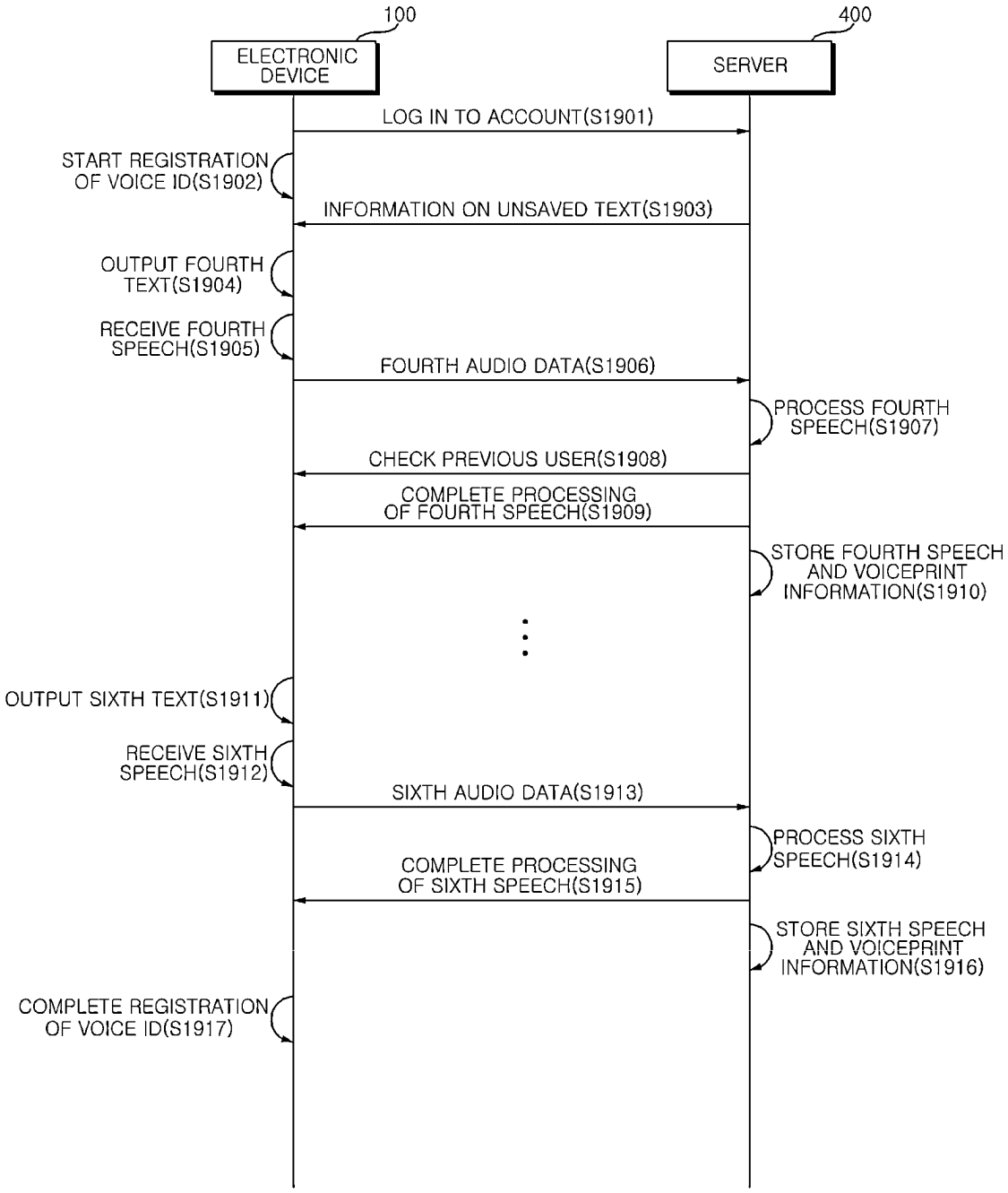
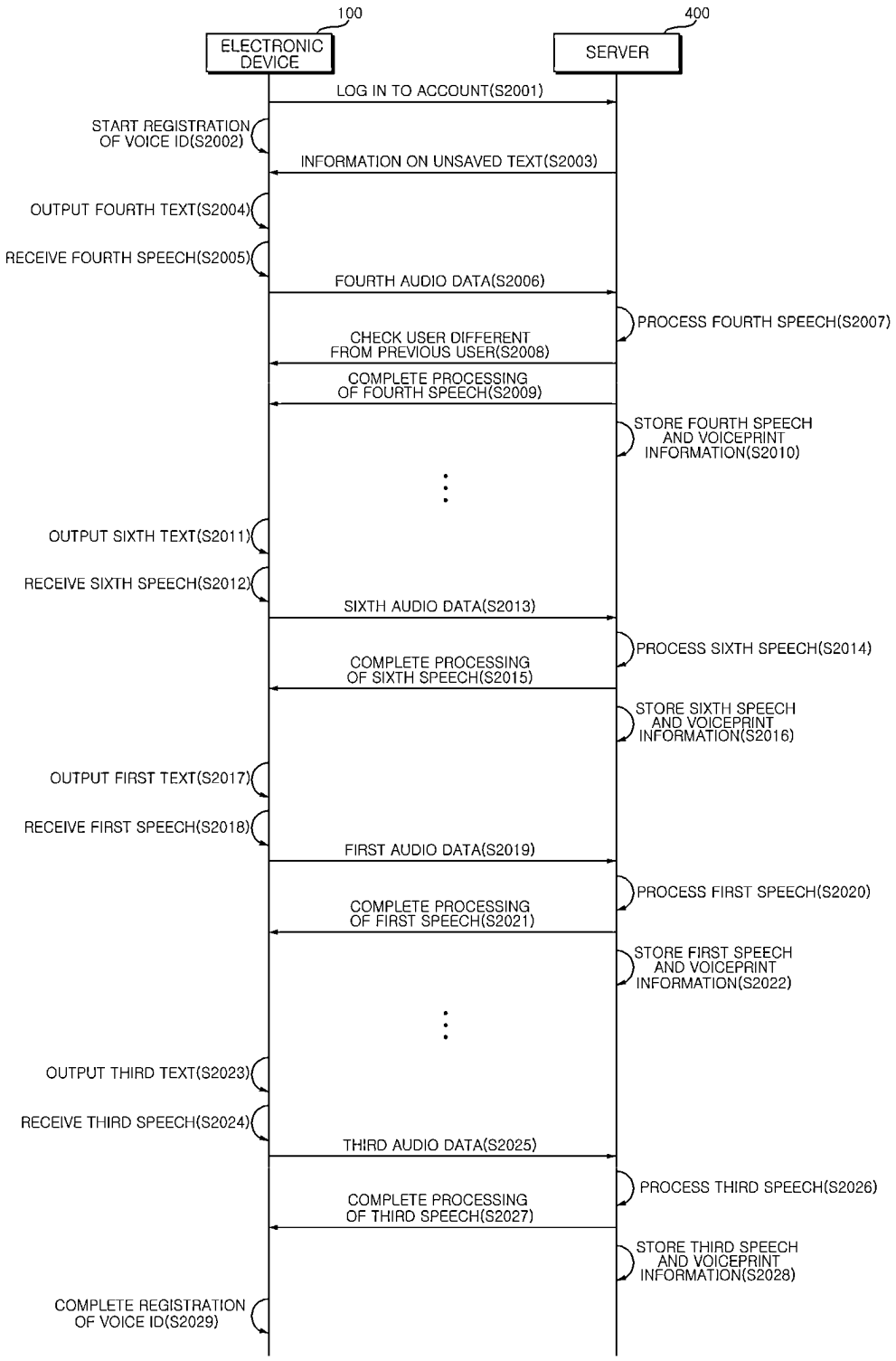

100
ELECTRONIC DEVICE
400
SERVER
LOG IN TO ACCOUNT(S1901)
START REGISTRATION OF VOICE ID(S1902)
INFORMATION ON UNSAVED TEXT(S1903)
OUTPUT FOURTH TEXT(S1904)
RECEIVE FOURTH SPEECH(S1905)
FOURTH AUDIO DATA(S1906)
PROCESS FOURTH SPEECH(S1907)
CHECK PREVIOUS USER(S1908)
COMPLETE PROCESSING OF FOURTH SPEECH(S1909)
STORE FOURTH SPEECH AND VOICEPRINT INFORMATION(S1910)
OUTPUT SIXTH TEXT(S1911)
RECEIVE SIXTH SPEECH(S1912)
SIXTH AUDIO DATA(S1913)
PROCESS SIXTH SPEECH(S1914)
COMPLETE PROCESSING OF SIXTH SPEECH(S1915)
STORE SIXTH SPEECH AND VOICEPRINT INFORMATION(S1916)
COMPLETE REGISTRATION OF VOICE ID(S1917)

FIG. 20


100
ELECTRONIC DEVICE
400
SERVER
LOG IN TO ACCOUNT(S2001)
START REGISTRATION OF VOICE ID(S2002)
INFORMATION ON UNSAVED TEXT(S2003)
OUTPUT FOURTH TEXT(S2004)
RECEIVE FOURTH SPEECH(S2005)
FOURTH AUDIO DATA(S2006)
PROCESS FOURTH SPEECH(S2007)
CHECK USER DIFFERENT FROM PREVIOUS USER(S2008)
COMPLETE PROCESSING OF FOURTH SPEECH(S2009)
STORE FOURTH SPEECH AND VOICEPRINT INFORMATION(S2010)
OUTPUT SIXTH TEXT(S2011)
RECEIVE SIXTH SPEECH(S2012)
SIXTH AUDIO DATA(S2013)
PROCESS SIXTH SPEECH(S2014)
COMPLETE PROCESSING OF SIXTH SPEECH(S2015)
STORE SIXTH SPEECH AND VOICEPRINT INFORMATION(S2016)
OUTPUT FIRST TEXT(S2017)
RECEIVE FIRST SPEECH(S2018)
FIRST AUDIO DATA(S2019)
PROCESS FIRST SPEECH(S2020)
COMPLETE PROCESSING OF FIRST SPEECH(S2021)
STORE FIRST SPEECH AND VOICEPRINT INFORMATION(S2022)
OUTPUT THIRD TEXT(S2023)
RECEIVE THIRD SPEECH(S2024)
THIRD AUDIO DATA(S2025)
PROCESS THIRD SPEECH(S2026)
COMPLETE PROCESSING OF THIRD SPEECH(S2027)
STORE THIRD SPEECH AND VOICEPRINT INFORMATION(S2028)
COMPLETE REGISTRATION OF VOICE ID(S2029)

ELECTRONIC DEVICE, SERVER, AND SYSTEM INCLUDING SAME

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0001] The present disclosure relates to an electronic device, a server, and a system including the same, and more specifically, to an electronic device, a server, and a system including the same that utilize speech recognition technology.

2. Description of the Related Art

[0002] With the recent development of technology, research on speech recognition technology for processing speech is being actively conducted. In particular, research on speech recognition technology, which began with smartphones, is being conducted widely in various fields related to user convenience, such as vehicles, as well as home appliances used at home and in offices.

[0003] Speech recognition technology is commonly used when a user controls an electronic device using his or her voice. For example, when a user utters a command to control an electronic device, the electronic device may directly recognize and process user's speech and operate according to the command related to the speech, or may send the speech to a server that processes speech and then operate according to □ a command related to the speech received from the server.

[0004] Meanwhile, services or functions provided through electronic devices are becoming increasingly diverse. Additionally, users register accounts for various services and then use the services by logging in with the registered account. In this case, service providers use user information managed for each account to provide optimal functions or information tailored to the user.

[0005] Conventionally, when attempting to log in to use a service, a user needs to directly input account information, for example, an identification (ID) and/or a password of the account. However, it is inconvenient for a user to input account information one by one into various services. Additionally, if a user remains logged in to eliminate the inconvenience of inputting account information, security problems such as access to the user's account information by others may occur. Additionally, when multiple users use one electronic device together, there is a problem that multiple users need to input their account information and log in each time they use a service.

SUMMARY OF THE INVENTION

[0006] Therefore, the present disclosure has been made in view of the above problems, and it is an object of the present disclosure to solve the above-described problems and other problems.

[0007] It is another object of the present disclosure to provide an electronic device, a server, and a system including the same capable of registering identification information on a user voice in a user account.

[0008] It is a further object of the present disclosure to provide an electronic device, a server, and a system including the same capable of identifying a user based on a user voice.

[0009] It is a further object of the present disclosure to provide an electronic device, a server, and a system including the same capable of logging in with a user account identified based on user voice.

[0010] It is a further object of the present disclosure to provide an electronic device, a server, and a system including the same capable of maintaining the continuity of identification information on a user who have previously attempted registration in a process of registering identification information on user voice to a user account.

[0011] In accordance with the present disclosure, the above and other objects can be accomplished by the provision of an electronic device including a display, an external device interface configured to communicate with a remote control device, a network interface configured to communicate with a server, a user input interface configured to transmit a signal related to a user input, and a controller, wherein the controller is configured to display a preset text related to registration of identification information through the display while performing a process of registering the identification information related to speech for a user account logged in to the server, upon receiving a speech signal related to the preset text through the user input interface, transmit data including the speech signal to the server, complete the process of registering the identification information based on processing of the speech signal related to the preset text by the server, and interrupt the process of registering the identification information upon receiving a predetermined input related to speech recognition from the remote control device during the process of registering the identification information.

[0012] In accordance with another aspect of the present disclosure, there is provided a server including a communication interface configured to communicate with an electronic device, a database, and a controller, wherein the controller is configured to convert a speech signal included in data received from the electronic device into text while performing a process of registering identification information related to speech for a user account logged in to the server, generate identification information related to the speech signal based on the converted text being related to preset text related to registration of the identification information, map the generated identification information to user identification information related to the user account and store the mapped information in the database, and maintain or delete the identification information mapped to the user identification information upon receiving a notification regarding interruption of the process of registering the identification information from the electronic device during the process of registering the identification information.

[0013] In accordance with another aspect of the present disclosure, there is provided a system including an electronic device, and a server, wherein the electronic device is configured to display a preset text related to registration of identification information while performing a process of registering the identification information related to speech for a user account logged in to the server, upon receiving a speech signal related to the preset text, transmit data including the speech signal to the server, complete the process of registering the identification information based on processing of the speech signal related to the preset text by the server, and stop the process of registering the identification information upon receiving a predetermined input related to speech recognition from a remote control device during the

process of registering the identification information, and the server is configured to convert the speech signal included in the data received from the electronic device into text, generate identification information related to the speech signal based on the converted text and the preset text being related to each other, map the generated identification information to user identification information related to the user account and store the mapped information in a database, and maintain or delete the identification information mapped to the user identification information upon receiving a notification regarding interruption of the process of registering the identification information from the electronic device during the process of registering the identification information.

[0014] Additional scope of applicability of the present disclosure will become apparent from the detailed description that follows. However, since various changes and modifications within the scope of the present disclosure may be clearly understood by those skilled in the art, the detailed description and specific embodiments such as preferred embodiments of the present disclosure should be understood as being given only as examples.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The above and other objects, features and other advantages of the present disclosure will be more clearly understood from the following detailed description taken in conjunction with the accompanying drawings, in which:

[0016] FIG. **1** is a diagram illustrating a system according to an embodiment of the present disclosure;

[0017] FIG. **2** is an internal block diagram of an electronic device of FIG. **1**;

[0018] FIG. **3** is a diagram referenced in description of a server of FIG. **1**;

[0019] FIG. **4** is a block diagram illustrating a configuration of a server according to an embodiment of the present disclosure;

[0020] FIG. **5** is a diagram illustrating an example of converting a speech signal into a power spectrum according to an embodiment of the present disclosure;

[0021] FIG. **6** is a block diagram illustrating the configuration of a controller for speech recognition and synthesis of an electronic device, according to an embodiment of the present disclosure;

[0022] FIG. **7** is a flowchart of a method of operating an electronic device according to an embodiment of the present disclosure;

[0023] FIG. **8** is a flowchart of a method of operating a system according to an embodiment of the present disclosure;

[0024] FIG. **9** to FIG. **16** are diagrams referenced in description of a process of registering identification information on user voice in a user account according to an embodiment of the present disclosure;

[0025] FIG. **17**A and FIG. **17**B are flowcharts of a method of operating an electronic device according to another embodiment of the present disclosure; and

[0026] FIG. **18** to FIG. **20** are flowcharts of methods of operating a system according to various embodiments of the present disclosure.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0027] Hereinafter, the present disclosure will be described in detail with reference to the attached drawings. In the drawings, parts not related to description are omitted in order to clearly and briefly describe the present disclosure, and identical or extremely similar parts are denoted by the same reference numerals throughout the specification.

[0028] The suffixes "module" and "part" for components used in the following description are simply given in consideration of the ease of writing this specification and do not have any particularly important meaning or role. Accordingly, the terms "module" and "part" may be used interchangeably.

[0029] In the present disclosure, it will be further understood that the term "comprise" or "include" specifies the presence of a stated feature, figure, step, operation, component, part or combination thereof, but does not preclude the presence or addition of one or more other features, figures, steps, operations, components, or combinations thereof.

[0030] Further, in this specification, the terms "first" and/or "second" are used to describe various components, but such components are not limited by these terms. The terms are used to discriminate one component from another component.

[0031] FIG. **1** is a diagram illustrating a system according to various embodiments of the present disclosure.

[0032] Referring to FIG. **1**, the system **10** may include an electronic device **100** and/or a server **400**.

[0033] The electronic device **100** may transmit/receive data to/from at least one server **400**. For example, the electronic device **100** may transmit/receive data to/from the at least one server **400** via a network **300** such as the Internet.

[0034] According to an embodiment, the at least one server **400** may include a server that performs speech recognition, a server that processes data using a super-giant artificial intelligence model, a server that provides content, and the like.

[0035] The electronic device **100** may include an image display device **100***a*, an air conditioner **100***b*, a refrigerator **100***c*, an air purifier **100***d*, a washing machine **100***e*, a vehicle **100***f*, and the like. Although the electronic device **100** is an image display device **100***a* in the present disclosure, the present disclosure is not limited thereto.

[0036] The image display device **100***a* may be a device that processes and outputs images. The image display device **100***a* is not particularly limited as long as it can output a screen related to video signals, such as a TV, a laptop computer, or a monitor.

[0037] The image display device **100***a* may receive a broadcast signal, process the same, and output a processed broadcast image. When the image display device **100***a* receives a broadcast signal, the image display device **100***a* may correspond to a broadcast reception device.

[0038] The image display device **100***a* may receive broadcast signals wirelessly through an antenna, or may receive broadcast signals through a cable. For example, the image display device **100***a* may receive terrestrial broadcast signals, satellite broadcast signals, cable broadcast signals, and Internet protocol Television (IPTV) broadcast signals.

[0039] FIG. **2** is an internal block diagram of the electronic device of FIG. **1**.

[0040] Referring to FIG. **2**, the electronic device **100** may include a broadcast receiver **105**, an external device interface **130**, a network interface **135**, a storage **140**, a user input interface **150**, an input part **160**, a controller **170**, a display **180**, an audio output part **185**, and/or a power supply **190**.

[0041] The broadcast receiver **105** may include a tuner **110** and a demodulator **120**.

[0042] Meanwhile, the electronic device **100** may include only the broadcast receiver **105** and the external device interface **130** among the broadcast receiver **105**, the external device interface **130**, and the network interface **135**. That is, the electronic device **100** may not include the network interface **135**.

[0043] The tuner **110** may select a broadcast signal related to a channel selected by a user or broadcast signals of all previously stored channels among broadcast signals received through an antenna (not shown) or a cable (not shown). The tuner **110** may convert the selected broadcast signals into intermediate frequency signals or baseband video or audio signals.

[0044] For example, if a selected broadcast signal is a digital broadcast signal, the tuner **110** may convert the selected broadcast signal into a digital IF signal (DIF), and if the selected broadcast signal is an analog broadcast signal, convert the same into an analog baseband video or audio signal (CVBS/SIF). That is, the tuner **110** may process digital broadcast signals or analog broadcast signals. The analog base band video or audio signal (CVBS/SIF) output from the tuner **110** may be directly input to the controller **170**.

[0045] Meanwhile, the tuner **110** may sequentially select broadcast signals of all of stored broadcast channels through a channel memory function among received broadcast signals and convert the same into intermediate frequency signals or baseband video or audio signals.

[0046] The tuner **110** may include a plurality of tuners in order to receive broadcast signals of a plurality of channels. Alternatively, a single tuner that simultaneously receives broadcast signals of a plurality of channels may also be adopted.

[0047] The demodulator **120** may receive a digital IF signal (DIF) converted by the tuner **110** and perform a demodulation operation.

[0048] The demodulator **120** may output a stream signal TS after performing demodulation and channel decoding. Here, the stream signal may be a multiplexed video signal, audio signal, or data signal.

[0049] The stream signal output from the demodulator **120** may be input to the controller **170**. After performing demultiplexing and video/audio signal processing, the controller **170** may output video through the display **180** and output audio through the audio output part **185**.

[0050] The external device interface **130** may transmit/receive data to/from a connected external device. To this end, the external device interface **130** may include an A/V input/output part (not shown).

[0051] The external device interface **130** may be connected to external devices such as a digital versatile disc (DVD) player, a Blu-ray player, a game console, a camera, a camcorder, a computer (laptop), a set-top box, and the like in wired/wireless manners, and may also perform input/output operations with respect to external devices.

[0052] In addition, the external device interface **130** may establish a communication network with respect to various remote control devices **200** to receive control signals related to the operation of the electronic device **100** from the remote control devices **200** or to transmit data related to the operation of the electronic device **100** to the remote control devices **200**.

[0053] The A/V input/output part may receive video and audio signals from an external device. For example, the A/V input/output part may include an Ethernet terminal, a USB terminal, a composite video banking Sync (CVBS) terminal, a component terminal, an S-video terminal (analog), a digital visual interface (DVI) terminal, a high definition multimedia interface (HDMI) terminal, a mobile high-definition link (MHL) terminal, an RGB terminal, a D-SUB terminal, an IEEE 1394 terminal, an SPDIF terminal, a liquid HD terminal, and the like. Digital signals input through these terminals may be transmitted to the controller **170**. Here, analog signals input through the CVBS terminal and the S-video terminal may be converted into digital signals through an analog-to-digital converter (not shown) and transmitted to the controller **170**.

[0054] The external device interface **130** may include a wireless communication part (not shown) for short-distance wireless communication with other electronic devices. The external device interface **130** may exchange data with a neighboring mobile terminal through the wireless communication part. For example, the external device interface **130** may receive device information, executing application information, application images, and the like from the mobile terminal in a mirroring mode.

[0055] The external device interface **130** may perform short-range wireless communication using Bluetooth, radio frequency identification (RFID), infrared data association (IrDA), ultra-wideband (UWB), ZigBee, and the like.

[0056] The network interface **135** may provide an interface for connecting the electronic device **100** to a wired/wireless network including the Internet.

[0057] The network interface **135** may include a communication module (not shown) for connection to a wired/wireless network. For example, the network interface **135** may include a communication module for a wireless LAN (WLAN) (Wi-Fi), wireless broadband (WiBro), world interoperability for microwave access (WiMax), and high speed downlink packet access (HSDPA).

[0058] The network interface **135** may transmit/receive data to/from other users or other electronic devices through a connected network or another network linked to the connected network.

[0059] The network interface **135** may receive web content or data provided by content providers or network operators. That is, the network interface **135** may receive content such as movies, advertisements, games, VOD, and broadcasting and information related thereto provided from content providers or network providers through networks.

[0060] The network interface **135** may receive firmware update information and update files provided by network operators, and may transmit data to the Internet, content providers, or network operators.

[0061] The network interface **135** may select and receive a desired application from among applications open to the public through a network.

[0062] The storage **140** may store programs for processing and controlling each signal in the controller **170** and may store processed video, audio, or data signals. For example, the storage **140** may store application programs designed for

the purpose of performing various tasks that may be processed by the controller **170** and selectively provide some of the stored application programs at the request of the controller **170**.

[0063] Programs stored in the storage **140** are not particularly limited as long as they can be executed by the controller **170**.

[0064] The storage **140** may execute a function of temporarily storing video, voice, or data signals received from an external device through the external device interface **130**.

[0065] The storage **140** may store information on a predetermined broadcast channel through a channel memory function such as a channel map.

[0066] Although FIG. **2** illustrates an embodiment in which the storage **140** is provided separately from the controller **170**, the scope of the present disclosure is not limited thereto, and the storage **140** may be included in the controller **170**.

[0067] The storage **140** may include at least one of a volatile memory (e.g., a DRAM, an SRAM, an SDRAM, etc.) or a non-volatile memory (e.g., a flash memory, a hard disk drive (HDD), a solid-state drive (SSD), etc.). In various embodiments of the present disclosure, "storage" and "memory" may be used interchangeably.

[0068] The user input interface **150** may transmit a signal input by a user to the controller **170** or transmit a signal from the controller **170** to the user.

[0069] For example, the user input interface **150** may transmit/receives user input signals such as power on/off, channel selection, and screen settings to/from the remote control device **200**, transmit user input signals input through local keys (not shown) such as a power key, a channel key, a volume key, and a setting key to the controller **170**, transmit a user input signal input through a sensor (not shown) that senses a user's gesture to the controller **170**, or transmit signals from the controller **170** to the sensor.

[0070] The input part **160** may be provided on one side of the main body of the electronic device **100**. For example, the input part **160** may include a touch pad, physical buttons, and the like.

[0071] The input part **160** may receive various user commands related to the operation of the electronic device **100** and transmit control signals related to the input commands to the controller **170**.

[0072] The input part **160** may include at least one microphone (not shown) and may receive a user voice through the microphone.

[0073] The controller **170** may include at least one processor and may control the overall operation of the electronic device **100** using the processor included therein. Here, the processor may be a general processor such as a central processing unit (CPU). The processor may be a dedicated device such as an ASIC or another hardware-based processor.

[0074] The controller **170** may demultiplex streams input through the tuner **110**, the demodulator **120**, the external device interface **130**, or the network interface **135**, or process demultiplexed signals to generate and output signals for video or audio output.

[0075] The display **180** may convert a video signal, a data signal, an OSD signal, and a control signal processed by the controller **170** or a video signal, a data signal, and a control signal received from the external device interface **130** to generate driving signals.

[0076] The display **180** may include a display panel (not shown) having a plurality of pixels.

[0077] The plurality of pixels provided in the display panel may include RGB subpixels. Alternatively, the plurality of pixels provided in the display panel may include RGBW subpixels. The display **180** may convert a video signal, a data signal, an OSD signal, a control signal, etc. processed by the controller **170** to generate driving signals for the plurality of pixels.

[0078] The display **180** may be a plasma display panel (PDP), a liquid crystal display (LCD), an organic light emitting diode (OLED) display, or a flexible display, and may also be a 3D display. 3D displays **180** may be classified into a glasses-free type and a glasses type.

[0079] Meanwhile, the display **180** may be configured as a touch screen and used as an input device in addition to an output device.

[0080] The audio output part **185** receives the audio signal processed by the controller **170** and outputs the same as audio.

[0081] A video signal processed by the controller **170** may be input to the display **180** and displayed as an image related to the video signal. Additionally, the video signal processed by the controller **170** may be input to an external output device through the external device interface **130**.

[0082] An audio signal processed by the controller **170** may be output as sound to the audio output part **185**. Additionally, the audio signal processed by the controller **170** may be input to an external output device through the external device interface **130**.

[0083] Although not illustrated in FIG. **2**, the controller **170** may include a demultiplexer, an image processor, etc.

[0084] In addition, the controller **170** may control overall operations of the electronic device **100**. For example, the controller **170** may control the tuner **110** to select (tune to) a broadcast related to a channel selected by the user or a previously stored channel.

[0085] Additionally, the controller **170** may control the electronic device **100** using a user command input through the user input interface **150** or an internal program.

[0086] Meanwhile, the controller **170** may control the display **180** to display an image. Here, the image displayed on the display **180** may be a still image or a video, and may be a 2D image or a 3D image.

[0087] Further, the controller **170** may cause a predetermined 2D object to be displayed in an image displayed on the display **180**. For example, the object may be at least one of a connected web screen (newspaper, magazine, or the like), an electronic program guide (EPG), various menus, widgets, icons, a still image, a video, or text.

[0088] Meanwhile, the electronic device **100** may further include an imaging device (not shown). The imaging device may capture an image of the user. The imaging device may be implemented as a single camera, but the present disclosure is not limited thereto and the imaging device may also be implemented as a plurality of cameras. Meanwhile, the imaging device may be embedded in the electronic device **100** at the top of the display **180** or may be disposed separately. Image information captured by the imaging device may be input to the controller **170**.

[0089] The controller **170** may recognize a location of the user based on images captured by the imaging device. For example, the controller **170** may ascertain the distance (z-axis coordinate) between the user and the electronic

device **100**. In addition, the controller **170** may ascertain the x-axis coordinate and y-axis coordinate in the display **180** related to the location of the user.

[0090] The controller **170** may detect a user's gesture based on images captured by the imaging device, each signal detected by a sensor, or a combination thereof.

[0091] The power supply **190** may supply corresponding power throughout the electronic device **100**. In particular, the power supply **190** may supply power to the controller **170**, which may be implemented in the form of a system on chip (SOC), the display **180** for displaying images, and the audio output part **185** for audio output.

[0092] Specifically, the power supply **190** may include a converter (not shown) that converts AC power to DC power and a DC/DC converter (not shown) that converts a DC power level.

[0093] The remote control device **200** may transmit user input to the user input interface **150**. To this end, the remote control device **200** may use Bluetooth, radio frequency (RF) communication, infrared communication, ultra-wideband (UWB), ZigBee, and the like. Additionally, the remote control device **200** may receive video, audio, or data signals output from the user input interface **150** and display the same or output the same as audio through the remote control device **200**.

[0094] The electronic device **100** described above may be a stationary or mobile digital broadcast receiver capable of receiving digital broadcasting.

[0095] Meanwhile, the block diagram of the electronic device **100** shown in FIG. **2** is merely a block diagram for an embodiment of the present disclosure, and components of the block diagram may be integrated, added, or omitted according to the specifications of the electronic device **100** that is actually implemented.

[0096] That is, two or more components may be combined into one component, or one component may be subdivided into two or more components as necessary. In addition, the function executed by each block is for describing an embodiment of the present disclosure, and the specific operation or device does not limit the scope of the present disclosure.

[0097] FIG. **3** is a diagram referenced in description of the server of FIG. **1**.

[0098] Referring to FIG. **3**, the server **400** may include a relay server **410**, a speech-to-text (STT) server **420**, a natural language processing (NLP) server **430**, a user identification server **440**, and/or an account server **450**. Although the relay server **410**, the STT server **420**, the NLP server **430**, the user identification server **440**, and the account server **450** are distinguished from each other in the present disclosure, the present disclosure is not limited thereto. For example, two or more of the relay server **410**, the STT server **420**, the NLP server **430**, the user identification server **440**, and the account server **450** may be configured as one server.

[0099] The relay server **410** may communicate with the electronic device **100**. The relay server **410** may transmit data between the STT server **420**, the NLP server **430**, the user identification server **440**, and the electronic device **100**. The relay server **410** may store at least some data transmitted between the STT server **420**, the NLP server **430**, the user identification server **440**, and the electronic device **100**.

[0100] The STT server **420** may receive audio data. The STT server **420** may convert the audio data into text data. The STT server **420** may transmit the text data to the electronic device **100** via the relay server **410**. The STT server **420** may be called an automatic speech recognition (ASR) server.

[0101] The STT server **420** may increase the accuracy of speech-to-text conversion using a language model. A language model may refer to a model that may calculate the probability of a sentence or the probability of the next word appearing when previous words are provided. For example, the language model may include probabilistic language models such as Unigram model, Bigram model, and N-gram model. That is, the STT server **420** may determine whether text data has been appropriately converted from audio data, and accordingly, increase the accuracy of conversion to text data.

[0102] The NLP server **430** may receive text data. The NLP server **430** may perform intent analysis on the text data based on the received text data. The NLP server **430** may transmit intent analysis information indicating the result of intent analysis to the electronic device **100** via the relay server **410**.

[0103] According to an embodiment, the NLP server **430** may generate intent analysis information by sequentially performing a morpheme analysis step, a syntax analysis step, a speech-act analysis step, a conversation processing step, and the like on text data. The morpheme analysis step is a step of classifying text data related to speech uttered by a user into morpheme units, which are the smallest units with meaning, and determining to what part of speech each classified morpheme corresponds. The syntax analysis step is a step of classifying text data into noun phrases, verb phrases, adjective phrases, and the like using the results of the morpheme analysis step and determining what kind of relationship is present between the classified phrases. Through the syntax analysis step, subjects, objects, and modifiers of speech uttered by a user may be determined. The speech-act analysis step is a step of analyzing the intention of speech uttered by a user using the results of the syntax analysis step. Specifically, the speech-act analysis step is a step of determining the intention of a sentence, such as whether a user is asking a question, making a request, or simply expressing an emotion. The conversation processing step is a step of determining whether to reply to user's utterance, respond thereto, or ask a question for additional information.

[0104] The user identification server **440** may receive audio data. The user identification server **440** may extract voice features based on the audio data. Here, the voice features may include the waveform of the voice, the frequency band of the voice, the power spectrum of the voice, and the like. Extraction of voice features will be described later with reference to FIG. **4** and FIG. **5**.

[0105] The user identification server **440** may obtain a voice feature vector from the voice features. The user identification server **440** may obtain the voice feature vector from the voice features based on a linear predictive coefficient, cepstrum, Mel frequency cepstral coefficient (MFCC), and filter bank energy.

[0106] The user identification server **440** may determine a similarity between a plurality of feature vectors. The user identification server **440** may determine the similarity between the plurality of feature vectors using cosine similarity, Euclidean similarity, or the like. Although an example of calculating a similarity between a first voice input and a second voice input based on cosine similarity will be

described in the present disclosure, the method of determining a similarity is not limited thereto. For example, a first vector related to first text and a second vector related to second text may be created. A cosine similarity between the first vector and the second vector may be calculated based on Formula 1 below.

$$\cos(\Theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \quad \text{[Formula 1]}$$

[0107] Here, A·B indicates the dot product of two vectors, an ||A|| and ||B|| indicate the magnitudes of the two vectors. That is, cosine similarity may be calculated by dividing the dot product of two vectors by the product of the magnitudes of the vectors. Cosine similarity may range from −1 to 1, and two vectors are determined to be similar as the cosine similarity therebetween is closer to 1.

[0108] The user identification server **440** may determine whether users who have uttered speech are the same based on the similarity between a plurality of feature vectors. For example, when a similarity between a first feature vector related to the first voice input and a second feature vector related to the second voice input is equal to or greater than a predetermined standard, the user identification server **440** may determine that the user who has uttered the first voice input and the user who has uttered the second voice input are the same.

[0109] According to an embodiment, the user identification server **440** may obtain a vector by processing a voice feature vector using an algorithm such as the Gaussian mixture model (GMM), supervector, i-vector, d-vector, x-vector, or the like. The user identification server **440** may determine whether users who have uttered voices are the same based on a similarity between a first vector obtained by processing a first feature vector and a second vector obtained by processing a second feature vector.

[0110] The user identification server **440** may store audio data. The user identification server **440** may store data on voiceprint (hereinafter, voiceprint information). Here, voiceprint information may include a voice feature vector and/or a vector obtained by processing the voice feature vector.

[0111] The user identification server **440** may store a voice database. The voice database regarding voices may include unique identification information related to the electronic device **100** (hereinafter referred to as device identification information), unique identification information related to a user account (hereinafter referred to as user identification information), voice data mapped to user identification information, voiceprint information mapped to user identification information.

[0112] The device identification information, user identification information, audio data, and voiceprint information included in the voice database may be stored in the user identification server **440** in association with one another. For example, at least one piece of device identification information, a plurality of pieces of audio data, and/or a plurality of pieces of voiceprint information may be mapped to user identification information. That is, it may be interpreted that device identification information, audio data, and voiceprint information are mapped to a user account and stored in the user identification server **440**. In the present disclosure, an example in which a plurality of pieces of audio data and a plurality of pieces of voiceprint information are all mapped to user identification information included in the voice database will be described.

[0113] The user identification server **440** may update voiceprint information included in a voice database based on audio data included in the voice database. For example, the user identification server **440** may generate voiceprint information related to audio data included in the voice database using an algorithm different from a previously used algorithm. Here, the user identification server **440** may change the voiceprint information included in the voice database to the newly generated voiceprint information.

[0114] The account server **450** may manage data regarding user accounts. The account server **450** may manage user account IDs, passwords, user identification information, device identification information mapped to user accounts, and whether or not users agree to terms and conditions related to various functions.

[0115] The account server **450** may store a database regarding user accounts. The database regarding user accounts may include user account IDs, passwords, user identification information, device identification information mapped to the user accounts, registration dates and times of the user accounts, whether or not users agree to terms and conditions related to various functions, and dates and times when users agree to the terms and conditions.

[0116] The account server **450** may communicate with the electronic device **100**. For example, the account server **450** may create and register a user account based on data from the electronic device **100**. For example, the account server **450** may approve login of a user account based on an ID and a password received from the electronic device **100**.

[0117] FIG. **4** is a block diagram for describing the configuration of the server according to an embodiment of the present disclosure.

[0118] Referring to FIG. **4**, the server **400** may include a preprocessor **460**, a controller **470**, a communication interface **480**, and/or a database **490**.

[0119] The preprocessor **460** may preprocess speech received through the communication interface **480** or speech stored in the database **490**.

[0120] The preprocessor **460** may be implemented as a separate chip from the controller **470** or may be implemented as a chip included in the controller **470**.

[0121] The preprocessor **460** may receive a voice signal (uttered by a user) and filter noise signals from the voice signal before converting the received voice signal into text data.

[0122] If the preprocessor **460** is provided in the electronic device **100**, the preprocessor **460** may recognize a startup word for activating speech recognition of the electronic device **100**. The preprocessor **460** may convert the startup word received through the user input interface **150** into text data, and if the converted text data is text data related to a pre-stored startup word, determine that the startup word is recognized.

[0123] The preprocessor **460** may convert the noise-removed voice signal into a power spectrum.

[0124] A power spectrum may be a parameter that indicates a frequency component included in a temporally varying waveform of a voice signal and the magnitude of the frequency component.

[0125] A power spectrum shows a distribution of squared amplitude values according to the frequency of the waveform of a voice signal. This will be described with reference to FIG. **5**.

[0126] FIG. **5** is a diagram illustrating an example of converting a voice signal into a power spectrum according to an embodiment of the present disclosure.

[0127] FIG. **5** shows a voice signal **510**. The voice signal **510** may be a signal received from an external device or may be a signal previously stored in the memory **170**.

[0128] The x-axis of the voice signal **510** represents time, and the y-axis represents amplitude.

[0129] A power spectrum processor **463** may convert the voice signal **510** in which the x-axis is the time axis into a power spectrum **520** in which the x-axis is the frequency axis. The power spectrum processor **463** may convert the voice signal **510** into the power spectrum **520** using Fast Fourier transform (FFT). The x-axis of the power spectrum **520** represents frequency, and the y-axis represents the square of amplitude.

[0130] Referring back to FIG. **4**, the functions of the preprocessor **460** and the controller **470** described in FIG. **4** may also be performed in the NLP server **430**.

[0131] The preprocessor **460** may include a wave processor **461**, a frequency processor **462**, the power spectrum processor **463**, a speech-to-text (STT) converter **464**, and the like.

[0132] The wave processor **461** may extract the waveform of speech.

[0133] The frequency processor **462** may extract the frequency band of the speech.

[0134] The power spectrum processor **463** may extract the power spectrum of the speech.

[0135] A power spectrum may be a parameter that indicates, when a temporally varying waveform is given, a frequency component included in the waveform and the magnitude of the frequency component.

[0136] The STT converter **464** may convert speech into text. The STT converter **464** may convert speech in a specific language into text in that language.

[0137] The controller **470** may control the overall operation of the server **400**. The controller **470** may include a speech analyzer **471**, a text analyzer **472**, a feature clustering part **473**, a text mapper **474**, and/or a speech synthesizer **475**.

[0138] The speech analyzer **471** may extract speech characteristic information using one or more of the waveform of speech, the frequency band of the speech, and the power spectrum of the speech preprocessed in the preprocessor **460**. The speech characteristic information may include one or more of information on the sex of a speaker, the voice (or tone) of the speaker, the pitch of voice, the speaking style of the speaker, the speech rate of the speaker, and the emotion of the speaker. Additionally, the speech characteristic information may further include the timbre of the speaker.

[0139] The text analyzer **472** may extract main expressions from text converted by the STT converter **464**. Upon detecting a change in tone between phrases from the converted text, the text analyzer **472** may extract the phrase with a different tone as a main expression phrase. The text analyzer **472** may determine that the tone has changed when the frequency band between phrases has changed more than a preset band. The text analyzer **472** may extract key words from phrases in the converted text. A key word may be a noun present in a phrase, but this is merely an example.

[0140] The feature clustering part **473** may classify the speech type of the speaker using the speech characteristic information extracted by the speech analyzer **471**. The feature clustering part **473** may classify the speech type of the speaker by assigning a weight to each type item constituting the speech characteristic information. The feature clustering part **473** may classify the speech type of the speaker using an attention technique of a deep learning model.

[0141] The text mapper **474** may translate text converted into a first language into text in a second language. The text mapper **474** may map the text translated into the second language with the text in the first language. The text mapper **474** may map main expressions constituting the text in the first language to corresponding phrases in the second language. The text mapper **474** may map a speech type related to the main expressions constituting the text in the first language to phrases in the second language. This is for the purpose of applying the classified speech type to the phrases in the second language.

[0142] The speech synthesizer **475** may apply the speech type and speaker's tone classified by the feature clustering part **473** to the main expressions of the text translated into the second language in the text mapper **474** to generate synthetic speech.

[0143] The controller **470** may determine the speech characteristics of the user using one or more of the transmitted text data or the power spectrum **520**.

[0144] Speech characteristics of a user may include the sex, pitch, tone, speech topic, speech rate, and voice volume of the user.

[0145] The controller **470** may obtain the frequency of the voice signal **510** and the amplitude corresponding to the frequency.

[0146] The controller **470** may determine the sex of the user who has uttered the voice using the frequency band of the power spectrum **470**. For example, if the frequency band of the power spectrum **520** is within a preset first frequency band range, the controller **470** may determine that the user is male.

[0147] If the frequency band of the power spectrum **520** is within a preset second frequency band range, the controller **470** may determine that the user is female. Here, the second frequency band range may be higher than the first frequency band range.

[0148] The controller **470** may determine the pitch of voice using the frequency band of the power spectrum **520**. For example, the controller **470** may determine the pitch of the voice based on the amplitude within a specific frequency band.

[0149] The controller **470** may determine the user's tone using the frequency band of the power spectrum **520**. For example, the controller **470** may determine a frequency band with an amplitude equal to or greater than a certain level among the frequency bands of the power spectrum **520** as a main sound range of the user and determine this main sound range as the user's tone.

[0150] The controller **470** may determine the user's speech rate based on the number of syllables uttered per unit time from the converted text data.

[0151] The controller **470** may determine the topic of the user's speech using the Bag-Of-Word Model technique for the converted text data.

[0152] The Bag-Of-Word Model technique is a technique of extracting frequently used words based on the frequency of a word in a sentence. Specifically, the Bag-Of-Word Model technique is a technique of extracting unique words within a sentence and expressing the frequency of each extracted word as a vector to determine the features of the topic of speech. For example, if words such as "running" and "physical strength" appear frequently in text data, the controller **470** may classify the topic of the user's speech as exercise.

[0153] The controller **470** may determine the topic of the user's speech from the text data using a known text categorization technique. The controller **470** may extract keywords from the text data and determine the topic of the user's speech.

[0154] The controller **470** may determine the user's voice volume by considering amplitude information in the entire frequency band. For example, the controller **470** may determine the user's voice volume based on the average or weighted average of amplitudes in each frequency band of the power spectrum **470**.

[0155] The communication interface **480** may communicate with an external server by wire or wirelessly. The communication interface **480** may communicate with the electronic device **100** by wire or wirelessly.

[0156] The database **490** may store speech in first language included in content. The database **490** may store synthetic speech in which speech in the first language has been converted into speech in the second language. The database **490** may store first text related to speech in the first language and second text in which the first text has been translated into the second language. The database **490** may store various learning models required for speech recognition.

[0157] Meanwhile, the controller **170** of the electronic device **100** illustrated in FIG. **2** may include the preprocessor **460** and the controller **470** illustrated in FIG. **4**. That is, the controller **170** of the electronic device **100** may perform the functions of the preprocessor **460** and the controller **470**.

[0158] FIG. **6** is a block diagram illustrating a configuration of a controller for speech recognition and synthesis of an image display device according to an embodiment of the present disclosure.

[0159] That is, the speech recognition and synthesis process illustrated in FIG. **6** may be performed by the controller **170** of the electronic device **100** without using the server.

[0160] Referring to FIG. **6**, the processor **170** of the electronic device **100** may include an STT engine **610**, an NLP engine **620**, and a speech synthesis engine **630**. Each engine may be either hardware or software.

[0161] The STT engine **610** may perform the function of the STT server **420** of FIG. **5**. That is, the STT engine **610** may convert audio data into text data.

[0162] The NLP engine **620** may perform the function of the NLP server **430** shown in FIG. **5**. That is, the NLP engine **620** may obtain intent analysis information indicating the speaker's intention from the converted text data.

[0163] The speech synthesis engine **630** may perform a function of a speech synthesis server. The speech synthesis engine **630** may search a database for syllables or words related to given text data and synthesize a combination of the searched syllables or words to generate synthetic speech.

[0164] The speech synthesis engine **630** may include a preprocessing engine **631** and a TTS engine **632**.

[0165] The preprocessing engine **631** may preprocess text data before generating synthetic speech. Specifically, the preprocessing engine **631** performs tokenization to divide text data into tokens, which are meaningful units. After performing tokenization, the preprocessing engine **631** may perform a cleansing operation to remove unnecessary characters and symbols to eliminate noise. Thereafter, the preprocessing engine **631** may generate the same word token by integrating word tokens with different expression methods. Thereafter, the preprocessing engine **631** may remove meaningless word tokens (stopwords).

[0166] The TTS engine **632** may synthesize speech related to the preprocessed text data and generate synthetic speech.

[0167] FIG. **7** is a flowchart of a method of operating an electronic device according to an embodiment of the present disclosure.

[0168] Referring to FIG. **7**, the electronic device **100** may determine whether a user account is logged in to the server **400** in operation S**701**. For example, a user may log in to the server **400** with a user account by entering the user account ID and password.

[0169] According to an embodiment, when the user first logs in to the server **400** using the electronic device **100** with the user account, the electronic device **100** may include user identification information related to the user account in a user list. For example, in a case where three different user accounts log in to the server **400** using the electronic device **100**, the user list stored in the electronic device **100** may include three different pieces of user identification information.

[0170] In operation S**702**, the electronic device **100** may determine whether voice-related identification information (hereinafter referred to as voice ID) is registered with respect to the user account logged in to the server **400**. Here, the voice ID may include voiceprint information stored in the user identification server **440**. For example, the server **400** may transmit information on whether a voice ID has been registered with respect to the user account logged in to the server **400** to the electronic device **100**.

[0171] According to an embodiment, the server **400** may determine whether the voice ID has been registered based on whether voiceprint information has been mapped to user identification information, which is unique identification information related to the user account logged in to the server **400**. Here, when the voice ID has not been registered with respect to the user account, the number of pieces of voiceprint information mapped to the user identification information may be 0.

[0172] According to an embodiment, the server **400** may determine that the voice ID has been registered if the number of pieces of voiceprint information mapped to the user identification information is two or more predetermined numbers and determine that the voice ID has not been registered if the number of pieces of voiceprint information is less than the predetermined numbers. For example, in the case of a user account for which a voice ID has been registered, six different pieces of voiceprint information may be mapped to user identification information. For example, in the case of a user account for which a voice ID has not been registered, five or fewer voiceprint information may be mapped to user identification information.

[0173] According to an embodiment, a flag value indicating whether a voice ID has been registered may be mapped to user identification information stored in the server **400**.

Here, user identification information to which a flag value is mapped may be stored in the user identification server **440** and/or the account server **450**. The server **400** may determine whether the voice ID has been registered based on the flag value mapped to the user identification information. For example, a flag value mapped to user identification information may be 0 in the case of a user account for which a voice ID has not been registered, and a flag value mapped to user identification information may be 1 in the case of a user account for which a voice ID has been registered.

[0174] When the voice ID has not been registered with respect to the user account, the electronic device **100** may start a process of registering the voice ID in operation S**703**. For example, when starting the process of registering the voice ID, the electronic device **100** may transmit data containing the device identification information, the user identification information, a value indicating the start of registration of the voice ID, etc. to the server **400**.

[0175] The electronic device **100** may output preset text in operation S**704**. The electronic device **100** may output any one of a plurality of pieces of preset text. For example, when the electronic device **100** is the image display device **100***a*, the electronic device **100** may output preset text through the display **180**.

[0176] According to an embodiment, the server **400** may transmit any one of a plurality of pieces of preset text to the electronic device **100** in a preset order. Here, the electronic device **100** may output the preset text received from the server **400**.

[0177] The electronic device **100** may determine whether speech with respect to the preset text is input in operation S**705**. For example, the electronic device **100** may determine whether speech is input through a microphone included in the input part **160** within a preset time. Here, the voice signal related to the speech input through the microphone may be transmitted to the controller **170** through the user input interface **150**. For example, the electronic device **100** may determine whether data containing a voice signal related to speech uttered by the user is received from the remote control device **200** within a preset time.

[0178] When speech with respect to the preset text is input, the electronic device **100** may transmit audio data including the voice signal related to the speech to the server **400** in operation S**706**. Here, the electronic device **100** may transmit the device identification information, the user identification information, and a language code indicating the type of language to the server **400** along with the audio data.

[0179] The server **400** may convert the voice signal included in the audio data received from the electronic device **100** into text. The server **400** may determine whether the text converted from the voice signal and the preset text correspond to each other. For example, the server **400** may determine whether the text converted from the voice signal and the preset text correspond to each other based on the similarity therebetween.

[0180] The server **400** may generate voiceprint information related to the voice signal when the text converted from the voice signal and the preset text correspond to each other. The server **400** may map the voiceprint information generated with respect to the preset text to the user identification information and store the same. The server **400** may map the audio data received with respect to the preset text to the user identification information and store the same.

[0181] The electronic device **100** may determine whether speech processing for the preset text is successful based on the response received from the server **400** in operation S**707**. For example, if the text converted from the voice signal and the preset text correspond to each other, the server **400** may notify the electronic device **100** of success of speech processing. For example, when the voiceprint information related to the voice signal has been generated, the server **400** may notify the electronic device **100** of success of speech processing.

[0182] Meanwhile, in operation S**708**, the electronic device **100** may determine whether the user reattempts to input speech when speech with respect to the preset text is not input or when speech processing for the preset text fails. For example, the electronic device **100** may reattempt to input speech based on a user input from the user reattempting to input speech. Here, the electronic device **100** may output the preset text again.

[0183] In operation S**709**, if speech processing for the preset text is successful, the electronic device **100** may determine whether processing for all pieces of text is completed. For example, if all speech processing for six pieces of text is successful, processing for all pieces of text may be completed. Meanwhile, when processing for five pieces of preset text is completed, the electronic device **100** may output the last preset text.

[0184] The electronic device **100** may end the process of registering the voice ID when processing for all pieces of text is completed in operation S**710**. For example, when the electronic device **100** is the image display device **100***a*, the electronic device **100** may output a screen indicating completion of voice ID registration through the display **180**. For example, the electronic device **100** may transmit data indicating completion of voice ID registration to the account server **450**.

[0185] According to one embodiment, the electronic device **100** may log in to the server **400** with a user account with a registered voice ID based on speech input to the electronic device **100**.

[0186] When speech is input, the electronic device **100** may transmit audio data related to the input speech to the server **400**. Here, the electronic device **100** may transmit device identification information, a user list, and a language code indicating the type of language to the server **400** along with the audio data.

[0187] The server **400** may generate voiceprint information regarding the speech input to the electronic device **100** based on the audio data received from the electronic device **100**. The server **400** may search an audio database for voiceprint information (hereinafter referred to as candidate voiceprint information) related to user identification information included in the user list received from the electronic device **100**. The server **400** may determine whether the candidate voiceprint information and the generated voiceprint information are related to each other. The server **400** may determine user identification information to which candidate voiceprint information related to the generated voiceprint information is mapped among the candidate voiceprint information as user identification information related to the speech input to the electronic device **100**. If there is no candidate voiceprint information related to the generated voiceprint information, the server **400** may determine that there is no user identification information related to the speech input to the electronic device **100**.

[0188] The server **400** may transmit a result of processing the audio data received from the electronic device **100** to the electronic device **100**. For example, the server **400** may transmit text converted from the audio data received from the electronic device **100**, a result of intent analysis performed on the converted text, user identification information related to the speech, etc. to the electronic device **100**.

[0189] The electronic device **100** may perform an operation related to the speech input to the electronic device **100** based on the audio data processing result received from the server **400**. For example, if the user identification information related to the speech is not related to the user account currently logged in to the server **400**, the electronic device **100** may log in to the server **400** with a user account related to the user identification information related to the speech. For example, if the user identification information related to the speech is related to the user account currently logged in to the server **400**, or if there is no user identification information related to the speech, the electronic device **100** may maintain the logged-in state of the user account currently logged in to the server **400**.

[0190] FIG. **8** is a flowchart of a method of operating a system according to an embodiment of the present disclosure.

[0191] Referring to FIG. **8**, the electronic device **100** may log in to the server **400** using a user account in operation S**801**.

[0192] The electronic device **100** may start a process of registering a voice ID in operation S**802**.

[0193] The electronic device **100** may output first text among a plurality of pieces of preset text in operation S**803**.

[0194] The electronic device **100** may receive first speech for the first text in operation S**804**.

[0195] The electronic device **100** may transmit first audio data including a speech signal related to the first speech to the server **400** in operation S**805**.

[0196] The server **400** may process the first speech for the first text based on the first audio data received from the electronic device **100** in operation S**806**. The server **400** may convert the speech signal related to the first speech included in the first audio data received from the electronic device **100** into text. The server **400** may determine whether the text converted from the speech signal related to the first speech and the first text correspond to each other.

[0197] The server **400** may notify the electronic device **100** of completion of processing for the first speech in operation S**807**. For example, the server **400** may notify the electronic device **100** of success of processing for the first speech based on the fact that the text converted from the speech signal related to the first speech corresponds to the first text.

[0198] Further, the server **400** may generate first voiceprint information with respect to the first speech based on the speech signal related to the first speech based on the fact that the text converted from the speech signal related to the first speech corresponds to the first text.

[0199] The server **400** may store the first audio data and the first voiceprint information with respect to the first speech in operation S**808**. The server **400** may map the first audio data and first voiceprint information to the user identification information related to the logged-in user account and store the same.

[0200] The electronic device **100** may output the second to fifth pieces of text in stages. The electronic device **100** may sequentially receive second to fifth speeches related to the second to fifth pieces of text. The electronic device **100** may sequentially transmit second to fifth pieces of audio data related to the second to fifth speeches to the server **400**.

[0201] The server **400** may process the second to fifth speeches based on the second to fifth pieces of audio data received from the electronic device **100**. Additionally, the server **400** may sequentially generate and store second to fifth pieces of speech information related to the second to fifth speeches.

[0202] The electronic device **100** may output sixth text from among a plurality of pieces of preset text in operation S**809**.

[0203] The electronic device **100** may receive sixth speech with respect to the sixth text in operation S**810**.

[0204] The electronic device **100** may transmit sixth audio data including a speech signal related to the sixth speech to the server **400** in operation S**811**.

[0205] The server **400** may process the sixth speech with respect to the sixth text based on the sixth audio data received from the electronic device **100** in operation S**812**. The server **400** may convert a speech signal related to the sixth speech included in the sixth audio data received from the electronic device **100** into text. The server **400** may determine whether the text converted from the speech signal related to the sixth speech and the sixth text correspond to each other.

[0206] The server **400** may notify the electronic device **100** of completion of processing for the sixth speech in operation S**813**.

[0207] Meanwhile, when the text converted from the speech signal related to the sixth speech and the sixth text correspond to each other, the server **400** may generate sixth voiceprint information regarding the sixth speech based on the speech signal related to the sixth speech.

[0208] The server **400** may store the sixth audio data and the sixth voiceprint information regarding the sixth speech in operation S**814**. The server **400** may map the sixth audio data and the sixth voiceprint information to the user identification information related to the logged-in user account and store the same. Here, six different pieces of audio data and a plurality of pieces of voiceprint information may be mapped to the user identification information related to the logged-in user account.

[0209] The electronic device **100** may end the process of registering the voice ID in operation S**815**. For example, the electronic device **100** may end the process of registering the voice ID based on completion of processing for the six different pieces of preset text.

[0210] Referring to FIG. **9**, if the user account is not logged in to the server **400**, the electronic device **100** may output a login screen **900** related to logging in to the server **400** through the display **180**. The login screen **900** may include an object **910** indicating a non-login state, and a login object **920** for executing login. When the user selects the login object **920** using a pointer **205** related to the remote control device **200**, the electronic device **100** may output a screen for entering an ID and a password. Here, the user may log in to the server **400** with the user account by entering the ID and the password of the user account.

[0211] Referring to FIG. **10**, when the voice ID has not been registered in the user account logged in to the server **400**, the electronic device **100** may output a first account screen **1000** related to the user account for which the voice

ID has not been registered. The first account screen **1000** may include an object **1010** indicating a logged-in user account, and an object **1020** regarding voice ID registration. When the user selects the object **1020** regarding voice ID registration using the pointer **205**, the electronic device **100** may start the process of registering a voice ID.

[0212] Referring to FIG. **11**A to FIG. **11**C, when a voice ID has been registered in a user account logged in to the server **400**, the electronic device **100** may display a second account screen **1100** related to the user account for which the voice ID has been registered. The second account screen **1100** may include an object **1110** indicating a logged-in user account, a re-registration object **1120** regarding voice ID re-registration, a deletion object **1130** regarding voice ID deletion, and an activation object **1140** regarding the use of a function related to voice ID. The user may select the activation object **1140** using the pointer **205** to activate or deactivate the use of a function related to voice ID.

[0213] When the user selects the reregistration object **1120** using the pointer **205**, the electronic device **100** may display a first notification screen **1121** for indicating that data related a previously registered voice ID is deleted from the server **400** due to reregistration of a voice ID. The user may select a confirmation object **1123** using the pointer **205** to reregister a voice ID, or select a cancellation object **1125** to maintain a previously registered voice ID.

[0214] When the user selects the deletion object **1130** using the pointer **205**, the electronic device **100** may display a second notification screen **1131** for indicating that data related to a previously registered voice ID is deleted from the server **400** due to deletion of a voice ID. The user may select a confirmation object **1133** using the pointer **205** to delete a voice ID, or select a cancellation object **1135** to maintain a previously registered voice ID.

[0215] Referring to FIG. **12**, when the object **1020** regarding voice ID registration is selected on the first account screen **1000**, or when a confirmation object **1123** is selected on a first notification screen **1121**, the electronic device **100** may output a start screen **1200** for starting voice ID registration. When the user selects a start object **1210** using the pointer **205**, the electronic device **100** may output a text screen for displaying preset text.

[0216] Referring to FIG. **13**, the electronic device **100** may output a text screen **1300** for displaying any one of a plurality of pieces of preset text. The text screen **1300** may include preset text **1301**, a text sequence number **1302**, an end object **1310** for ending the process of registering a voice ID, and an input object **1320** for receiving speech.

[0217] When the user selects the end object **1310** using the pointer **205**, the process of registering a voice ID may end. For example, when the process of registering a voice ID ends, all data stored in the server **400** while the process of registering a voice ID is in progress may be deleted.

[0218] When the user selects the input object **1320** using the pointer **205**, the electronic device **100** may receive speech with respect to text.

[0219] According to an embodiment, when the user presses a predetermined button (e.g., a voice input button) included in the remote control device **200** while the text screen **1300** is displayed, the electronic device **100** may receive speech with respect to text based on the user input of pressing the predetermined button, received from the remote control device **200**.

[0220] Meanwhile, according to an embodiment, when the user presses a predetermined button (e.g., the voice input button) included in the remote control device **200** while the process of registering a voice ID is in progress, the electronic device **100** may stop the process of registering a voice ID based on the user input of pressing the predetermined button, received from the remote control device **200**. Here, the user input of pressing a predetermined button (e.g., the voice input button) included in the remote control device **200** may correspond to a user input of starting speech recognition for speech received through the remote control device **200**. The electronic device **100** may perform an operation related to speech recognition on audio data including a speech signal received from the remote control device **200**.

[0221] Referring to FIG. **14**, when processing of speech with respect to text is successful, the electronic device **100** may display a success screen **1400** for indicating the success of the text processing. The success screen **1400** may include an object **1410** indicating the success of processing the speech with respect to the text.

[0222] Referring to FIG. **15**, when speech with respect to the text is not input or processing of the speech with respect to the text fails, the electronic device **100** may display a failure screen **1500** for indicating the failure of processing the speech with respect to the text. The failure screen **1500** may include a termination object **1510** indicating termination of the process of registering a voice ID, a re-input object **1520** indicating reattempt to receive speech, etc. When the user selects the re-input object **1520** using the pointer **205**, the electronic device **100** may receive speech with respect to the text again.

[0223] Referring to FIG. **16**, upon completion of processing of all pieces of text, the electronic device **100** may display a completion screen **1600** for indicating completion of voice ID registration. The completion screen **1600** may include an object **1610** indicating the user account for which the voice ID has been registered, a completion object **1620** indicating completion of the process of registering the voice ID, etc. When the user selects the completion object **1620** using the pointer **205**, the electronic device **100** may complete the process of registering the voice ID.

[0224] FIG. **17**A and FIG. **17**B are flowcharts of a method of operating an electronic device according to another embodiment of the present disclosure. Detailed description of redundant parts described in FIG. **7** will be omitted.

[0225] Referring to FIG. **17**A, the electronic device **100** may determine whether a user account is logged in to the server **400** in operation S**1701**.

[0226] The electronic device **100** may determine whether a voice ID has been registered for the user account logged in to the server **400** in operation S**1702**.

[0227] If no voice ID has been registered for the user account, the electronic device **100** may initiate a process of registering a voice ID in operation S**1703**.

[0228] The electronic device **100** may determine whether the user account is a user account for which a voice ID has been temporarily stored in operation S**1704**. For example, the server **400** may transmit information on whether a voice ID has been temporarily stored for the user account logged in to the server **400** to the electronic device **100**.

[0229] According to one embodiment, the server **400** may determine whether a voice ID has been temporarily stored based on whether voiceprint information is mapped to user

identification information, which is unique identification information related to the user account logged in to the server **400**. In the case of a user account for which a voice ID has been temporarily stored, more than a minimum number of pieces of voiceprint information may be mapped to the user identification information. Here, the minimum number may be less than a predetermined number of 6.

[0230] According to one embodiment, the server **400** may determine whether a voice ID has been temporarily stored based on a flag value mapped to user identification information. For example, in the case of a user account for which a voice ID has been temporarily stored, the flag value mapped to the user identification information may be 2.

[0231] The electronic device **100** may output preset text in operation S**1705**. For example, when no voice ID has been temporarily stored, the server **400** may transmit first text among a plurality of pieces of preset text to the electronic device **100**.

[0232] The electronic device **100** may determine whether speech with respect to the preset text is input in operation S**1706**.

[0233] When speech with respect to the preset text is input, the electronic device **100** may transmit audio data including a speech signal related to the speech to the server **400** in operation S**1707**.

[0234] The electronic device **100** may determine whether processing of the speech with respect to the preset text was successful based on a response received from the server **400** in operation S**1708**.

[0235] In operation S**1709**, the electronic device **100** may determine whether to reattempt speech input when speech with respect to the preset text is not input or processing of the speech with respect to the preset text fails.

[0236] In operation S**1710**, the electronic device **100** may determine whether processing of all pieces of text is completed when processing of the speech with respect to the preset text is successful.

[0237] The electronic device **100** may complete the process of registering a voice ID when processing of all pieces of text is completed in operation S**1711**.

[0238] Referring to FIG. **17**B, when a voice ID has been temporarily stored in operation S**1712**, the electronic device **100** may output any one (hereinafter, unsaved text) except text related to the voiceprint information mapped to the user identification information among the plurality of pieces of preset text. For example, when voiceprint information related to first text to third text is mapped to user identification information, which is unique identification information related to the user account logged in to the server **400**, the server **400** may transmit fourth text to the electronic device **100** as unsaved text. Here, the electronic device **100** may output the fourth text received from the server **400**.

[0239] The electronic device **100** may determine whether speech with respect to unsaved text is input in operation S**1713**.

[0240] When speech with respect to the unsaved text is input, the electronic device **100** may transmit audio data including a speech signal related to the speech to the server **400** in operation S**1714**.

[0241] The electronic device **100** may determine whether a previous user who has temporarily stored the voice ID is the same as the current user who has uttered the speech with respect to the unsaved text in operation S**1715**. For example, the electronic device **100** may determine whether the previous user is the same as the current user based on a result of determining whether the users are the same received from the server **400**.

[0242] The server **400** may convert the speech signal included in the audio data received from the electronic device **100** into text. The server **400** may determine whether the text converted from the speech signal is related to the unsaved text. For example, the server **400** may determine whether the converted text and the unsaved text are related to each other based on the similarity between the text converted from the speech signal and the unsaved text.

[0243] When the text converted from the speech signal and the unsaved text are related to each other, the server **400** may generate voiceprint information related to the speech with respect to the unsaved text (hereinafter, unsaved voiceprint information) based on the speech signal included in the audio data received from the electronic device **100**. The server **400** may determine whether at least one of voiceprint information mapped to user identification information is related to the unsaved voiceprint information. For example, when voiceprint information related to the first text to the third text is mapped to user identification information, the server **400** may calculate the similarity between first voiceprint information related to the first text and the unsaved voiceprint information. Here, the server **400** may determine that the first voiceprint information and the unsaved voiceprint information are related to each other when the similarity between the first voiceprint information and the unsaved voiceprint information is greater than a predetermined criterion. Additionally, the server **400** may determine that the previous user is the same as the current user when the first voiceprint information and the unsaved voiceprint information are related to each other.

[0244] If the previous user is different from the current user, the electronic device **100** may determine not to use data related to the previous user stored in the server **400** in operation S**1716**. For example, if the electronic device **100** is an image display device **100***a*, the electronic device **100** may display a screen for representing that the previous user is different from the current user through the display **180**.

[0245] The server **400** may maintain voiceprint information mapped to the user identification information when the previous user is the same as the current user. For example, in a state where the voiceprint information related to the first text to the third text is mapped to the user identification information, the server **400** may map the unsaved voiceprint information to the user identification information and store the same as voiceprint information related to the fourth text. Additionally, the server **400** may sequentially transmit fifth text and sixth text to the electronic device **100**.

[0246] The server **400** may delete the voiceprint information mapped to the user identification information when the previous user is different from the current user. Here, the server **400** may also delete audio data mapped to the user identification information. For example, the server **400** may delete the voiceprint information related to the first text to the third text mapped to the user identification information, and map the unsaved voiceprint information to the user identification information and store the same as voiceprint information related to the fourth text. Additionally, the server **400** may sequentially transmit text other than the fourth text among the plurality of pieces of preset text to the electronic device **100**.

[0247] In operation S**1717**, the electronic device **100** may determine whether to reattempt to input speech when no speech with respect to the unsaved text is input or processing of speech with respect to the unsaved text fails. For example, the electronic device **100** may reattempt to input speech based on user input for reattempting to input speech. Here, the electronic device **100** may output the unsaved text again.

[0248] FIG. **18** is a flowchart of a method of operating a system when a voice ID is not registered and temporarily stored according to an embodiment of the present disclosure.

[0249] Referring to FIG. **18**, the electronic device **100** may log in to the server **400** using a user account in operation S**1801**.

[0250] The electronic device **100** may initiate the process of registering a voice ID in operation S**1802**.

[0251] The electronic device **100** may output first text from among a plurality of pieces of preset text in operation S**1803**.

[0252] The electronic device **100** may receive first speech with respect to the first text in operation S**1804**.

[0253] The electronic device **100** may transmit first audio data including a speech signal related to the first speech to the server **400** in operation S**1805**.

[0254] The server **400** may process the first speech with respect to the first text based on the first audio data received from the electronic device **100** in operation S**1806**.

[0255] The server **400** may notify the electronic device **100** of completion of processing of the first speech in operation S**1807**.

[0256] The server **400** may store the first audio data and first voiceprint information on the first speech in operation S**1808**.

[0257] The electronic device **100** may output second text among the plurality of pieces of preset text in operation S**1809**.

[0258] The electronic device **100** may receive second speech with respect to the second text in operation S**1810**.

[0259] The electronic device **100** may transmit second audio data including a speech signal related to the second speech to the server **400** in operation S**1811**.

[0260] The server **400** may process the second speech with respect to the second text based on the second audio data received from the electronic device **100** in operation S**1812**.

[0261] The server **400** may notify the electronic device **100** of completion of processing of the second speech in operation S**1813**.

[0262] The server **400** may store the second audio data and second voiceprint information on the second speech in operation S**1814**.

[0263] The electronic device **100** may notify the server **400** of interruption of voice ID registration in operation S**1815**.

[0264] For example, when the user presses a power button included in the remote control device **200**, the electronic device **100** may be turned off based on a signal related to the user input of pressing the power button received from the remote control device **200**. Here, the electronic device **100** may notify the server **400** of interruption of voice ID registration based on the power being turned off.

[0265] For example, when the user presses a button related to a predetermined function (e.g., an OTT service button) included in the remote control device **200**, the electronic device **100** may execute the predetermined function based on a signal related to the user input of pressing the button related to the predetermined function (e.g., OTT service button) received from the remote control device **200**. In this case, the electronic device **100** may notify the server **400** of interruption of voice ID registration based on execution of the predetermined function. When execution of the predetermined function is terminated, the electronic device **100** may output a notification for notifying the user of interruption of voice ID registration. The electronic device **100** may determine whether to resume interrupted voice ID registration based on user input. When resuming interrupted voice ID registration, the electronic device **100** may initiate the process of registering a voice ID.

[0266] For example, when the user presses a speech recognition-related button (e.g., speech input button) included in the remote control device **200**, the electronic device **100** may execute a speech recognition function based on a signal related to the user input of pressing the speech recognition-related button received from the remote control device **200**. In this case, the electronic device **100** may notify the server **400** of interruption of voice ID registration based on execution of the speech recognition function. When execution of the speech recognition function is terminated, the electronic device **100** may output a notification for notifying the user of interruption of voice ID registration. The electronic device **100** may determine whether to resume interrupted voice ID registration based on user input. When resuming interrupted voice ID registration, the electronic device **100** may initiate the process of registering a voice ID.

[0267] In operation S**1816**, the server **400** may check the number of pieces of voiceprint information stored by being mapped to user identification information upon notification of interruption of voice ID registration from the electronic device **100**. Here, the server **400** may delete audio data and voiceprint information stored by being mapped to the user identification information based on the number of pieces of voiceprint information stored by being mapped to the user identification information being 2, which is less than a minimum number (e.g., 3).

[0268] Meanwhile, the server **400** may maintain the audio data and voiceprint information stored by being mapped to the user identification information based on the number of pieces of voiceprint information stored by being mapped to the user identification information being equal to or greater than the minimum number (e.g., 3).

[0269] FIG. **19** is a flowchart of a method of operating the system when a voice ID has been temporarily stored according to an embodiment of the present disclosure.

[0270] Referring to FIG. **19**, the electronic device **100** may log in to the server **400** using a user account in operation S**1901**.

[0271] The electronic device **100** may initiate the process of registering a voice ID in operation S**1902**.

[0272] The server **400** may transmit information on unsaved text to the electronic device **100** based on a voice ID having been temporarily stored in operation S**1903**. For example, the server **400** may transmit fourth text to the electronic device **100** as unsaved text based on voiceprint information related to first text to third text being mapped to user identification information, which is unique identification information related to the logged-in user account.

[0273] The electronic device **100** may output the fourth text, which is unsaved text, in operation S**1904**.

[0274] The electronic device **100** may receive fourth speech with respect to the fourth text in operation S**1905**.

[0275] The electronic device **100** may transmit fourth audio data including a speech signal related to the fourth speech to the server **400** in operation S**1906**.

[0276] The server **400** may process the fourth speech with respect to the fourth text based on the fourth audio data received from the electronic device **100** in operation S**1907**.

[0277] The server **400** may determine whether fourth voiceprint information generated based on the fourth audio data is related to at least one piece of voiceprint information mapped to the user identification information. For example, the server **400** may determine whether first voiceprint information related to the first text is related to the fourth voiceprint information based on the similarity between the first voiceprint information and the fourth voiceprint information.

[0278] In operation S**1908**, the server **400** may notify the electronic device **100** of the result of determining that a previous user who has temporarily stored the voice ID is the same as the user who has uttered the fourth speech. For example, the server **400** may determine that the previous user is the same as the current user based on the first voiceprint information and the fourth voiceprint information being related to each other.

[0279] The server **400** may notify the electronic device **100** of completion of processing of the fourth speech in operation S**1909**.

[0280] The server **400** may store the fourth audio data and fourth voiceprint information on the fourth speech in operation S**1910**. The server **400** may store the fourth audio data and the fourth voiceprint information by mapping the same to user identification information related to the logged-in user account.

[0281] The electronic device **100** may output fifth text. The electronic device **100** may receive fifth speech related to the fifth text. The electronic device **100** may transmit fifth audio data related to the fifth speech to the server **400**.

[0282] The server **400** may process the fifth speech based on the fifth audio data received from the electronic device **100**. Additionally, the server **400** may generate and store fifth audio information related to the fifth speech.

[0283] The electronic device **100** may output sixth text among the plurality of pieces of preset text in operation S**1911**.

[0284] The electronic device **100** may receive sixth speech with respect to the sixth text in operation S**1912**.

[0285] The electronic device **100** may transmit sixth audio data including a speech signal related to the sixth speech to the server **400** in operation S**1913**.

[0286] The server **400** may process the sixth speech with respect to the sixth text based on the sixth audio data received from the electronic device **100** in operation S**1914**.

[0287] The server **400** may notify the electronic device **100** of completion of processing of the sixth speech in operation S**1915**.

[0288] The server **400** may store the sixth audio data and sixth voiceprint information on the sixth speech in operation S**1916**.

[0289] The electronic device **100** may complete the process of registering a voice ID in operation S**1917**.

[0290] FIG. **20** is a flowchart of a method of operating the system when a voice ID is temporarily stored according to another embodiment of the present disclosure.

[0291] Referring to FIG. **20**, the electronic device **100** may log in to the server **400** using a user account in operation S**2001**.

[0292] The electronic device **100** may initiate the process of registering a voice ID in operation S**2002**.

[0293] The server **400** may transmit information on unsaved text to the electronic device **100** based on a voice ID having been temporarily stored in operation S**2003**. For example, the server **400** may transmit the fourth text to the electronic device **100** as unsaved text based on voiceprint information related to the first text to the third text being mapped to user identification information, which is unique identification information related to the logged-in user account.

[0294] The electronic device **100** may output the fourth text, which is unsaved text, in operation S**2004**.

[0295] The electronic device **100** may receive fourth speech with respect to the fourth text in operation S**2005**.

[0296] The electronic device **100** may transmit fourth audio data including a speech signal related to the fourth speech to the server **400** in operation S**2006**.

[0297] The server **400** may process the fourth speech with respect to the fourth text based on the fourth audio data received from the electronic device **100** in operation S**2007**.

[0298] The server **400** may notify the electronic device **100** of the result of determining that the previous user who has temporarily stored the voice ID and the user who has uttered the fourth speech are different from each other in operation S**2008**. For example, the server **400** may determine that the previous user and the current user are different from each other based on the first voiceprint information and the fourth voiceprint information not being related to each other. In this case, the server **400** may delete voiceprint information mapped to the identification information, which is unique identification information related to the logged-in user account.

[0299] The server **400** may notify the electronic device **100** of completion of processing of the fourth speech in operation S**2009**.

[0300] The server **400** may store the fourth audio data and the fourth voiceprint information on the fourth speech in operation S**2010**. The server **400** may store the fourth audio data and the fourth voiceprint information by mapping the same to the user identification information related to the logged-in user account.

[0301] Further, the server **400** may transmit any one of the preset text other than the fourth text to the electronic device **100**. In the present disclosure, an example in which the fifth text and the sixth text are sequentially transmitted to the electronic device **100** and then the first text to the third text are transmitted to the electronic device **100** will be described.

[0302] The electronic device **100** may output the fifth text. The electronic device **100** may receive fifth speech related to the fifth text. The electronic device **100** may transmit fifth audio data related to the fifth speech to the server **400**.

[0303] The server **400** may process the fifth speech based on the fifth audio data received from the electronic device **100**. Additionally, the server **400** may generate and store fifth audio information related to the fifth speech.

[0304] The electronic device **100** may output the sixth text among the plurality of pieces of preset text in operation S**2011**.

[0305] The electronic device **100** may receive sixth speech with respect to the sixth text in operation S**2012**.

[0306] The electronic device **100** may transmit sixth audio data including a speech signal related to the sixth speech to the server **400** in operation S**2013**.

[0307] The server **400** may process the sixth speech with respect to the sixth text based on the sixth audio data received from the electronic device **100** in operation S**2014**.

[0308] The server **400** may notify the electronic device **100** of completion of processing of the sixth speech in operation S**2015**.

[0309] The server **400** may store the sixth audio data and sixth voiceprint information on the sixth speech in operation S**2016**.

[0310] The electronic device **100** may output the first text among the plurality of pieces of preset texts in operation S**2017**.

[0311] The electronic device **100** may receive first speech with respect to the first text in operation S**2018**.

[0312] The electronic device **100** may transmit first audio data including a speech signal related to the first speech to the server **400** in operation S**2019**.

[0313] The server **400** may process the first speech with respect to the first text based on the first audio data received from the electronic device **100** in operation S**2020**.

[0314] The server **400** may notify the electronic device **100** of completion of processing of the first speech in operation S**2021**.

[0315] The server **400** may store the first audio data and first voiceprint information on the first speech in operation S**2022**.

[0316] The electronic device **100** may output the second text. The electronic device **100** may receive second speech with respect to the second text. The electronic device **100** may transmit second audio data related to the second speech to the server **400**.

[0317] The server **400** may process the second speech based on the second audio data received from the electronic device **100**. Additionally, the server **400** may generate and store second audio information related to the second speech.

[0318] The electronic device **100** may output the third text among the plurality of pieces of preset text in operation S**2023**.

[0319] The electronic device **100** may receive third speech with respect to the third text in operation S**2024**.

[0320] The electronic device **100** may transmit third audio data including a speech signal related to the third speech to the server **400** in operation S**2025**.

[0321] The server **400** may process the third speech with respect to the third text based on the third audio data received from the electronic device **100** in operation S**2026**.

[0322] The server **400** may notify the electronic device **100** of completion of processing of the third speech in operation S**2027**.

[0323] The server **400** may store the third audio data and third voiceprint information on the third speech in operation S**2028**.

[0324] The electronic device **100** may complete the process of registering a voice ID in operation S**2029**.

[0325] As described above, according to at least one embodiment of the present disclosure, it is possible to register identification information on user voice in a user account.

[0326] Additionally, according to at least one embodiment of the present disclosure, it is possible to identify a user based on user voice.

[0327] Additionally, according to at least one embodiment of the present disclosure, it is possible to log in to an account of a user identified based on user voice.

[0328] Additionally, according to at least one embodiment of the present disclosure, it is possible to maintain the continuity of identification information on a user who have previously attempted registration in a process of registering identification information on user voice to a user account.

[0329] Referring to FIG. **1** to FIG. **20**, the electronic device **100** according to one aspect of the present disclosure includes the display **180**, the external device interface **130** configured to communicate with remote control device **200**, the network interface **135** configured to communicate with the server **400**, the user input interface **150** configured to transmit a signal related to a user input, and the controller **170**, wherein the controller **170** is configured to display a preset text related to registration of identification information through the display **180** while performing a process of registering the identification information related to speech for a user account logged in to the server **400**, upon receiving a speech signal related to the preset text through the user input interface **150**, transmit data including the speech signal to the server **400**, complete the process of registering the identification information based on processing of the speech signal related to the preset text by the server **400**, and interrupt the process of registering the identification information upon receiving a predetermined input related to speech recognition from the remote control device **200** during the process of registering the identification information.

[0330] Additionally, according to one aspect of the present disclosure, the identification information may include a feature vector with respect to a voiceprint of the speech.

[0331] Additionally, according to one aspect of the present disclosure, the controller **170** may transmit, to the server **400**, data including a speech signal related to the predetermined input and received from the remote control device **200** after receiving the predetermined input, and perform an operation related to the predetermined input based on processing of the speech signal related to the predetermined input by the server **400**.

[0332] Additionally, according to one aspect of the present disclosure, the controller **170** may determine whether the identification information has been temporarily stored in the server **400** upon starting the process of registering the identification information for the user account, display a predetermined number of pieces of text in stages based on the identification information having not been temporarily stored in the server **400**, and display some of the plurality of pieces of text in stages excluding text related to the identification information temporarily stored in the server **400** based on the identification information having been temporarily stored in the server **100**.

[0333] Additionally, according to one aspect of the present disclosure, the controller **170** may display a notification regarding interruption of the process of registering the identification information through the display **180** based on termination of the operation related to the predetermined input, and determine whether to start the process of registering the identification information based on the user input received through the user input interface **150**.

[0334] The server **400** according to one aspect of the present disclosure includes the communication interface **480** configured to communicate with the electronic device **100**, the database **490**; and the controller **470**, wherein the controller **470** is configured to convert a speech signal included in data received from the electronic device **100** into text while performing a process of registering identification information related to speech for a user account logged in to the server **400**, generate identification information related to the speech signal based on the converted text being related to preset text related to registration of the identification information, map the generated identification information to user identification information related to the user account and store the mapped information in the database **490**, and maintain or delete the identification information mapped to the user identification information upon receiving a notification regarding interruption of the process of registering the identification information from the electronic device **100** during the process of registering the identification information.

[0335] Additionally, according to one aspect of the present disclosure, the identification information may include a feature vector with respect to a voiceprint of the speech.

[0336] Additionally, according to one aspect of the present disclosure, the controller **470** may maintain the identification information mapped to the user identification information based on a number of pieces of identification information mapped to the user identification information being equal to or greater than a preset minimum number, and delete the identification information mapped to the user identification information based on the number of pieces of identification information mapped to the user identification information being less than the preset minimum number.

[0337] Additionally, according to one aspect of the present disclosure, the controller **470** may determine whether the identification information is mapped to the user identification information based on the process of registering the identification information being started for the user account, transmit a predetermined number of pieces of text to the electronic device **100** in stages based on the identification information not being mapped to the user identification information, and transmit some of the plurality of pieces of text excluding text related to the identification information mapped to the user identification information to the electronic device **100** in stages based on the identification information being mapped to the user identification information.

[0338] Additionally, according to one aspect of the present disclosure, the controller **470** may determine whether the identification information is mapped to the user identification information based on the process of registering the identification information being started for the user account, transmit first text excluding the text related to the identification information mapped to the user identification information among the plurality of pieces of text to the electronic device **100** based on the identification information being mapped to the user identification information, generate identification information on a first speech signal related to the first text and included in data received from the electronic device **100**, maintain the identification information mapped to the user identification information based on the identification information on the first speech signal being related to at least one piece of the identification information mapped to the user identification information, and delete the identification information mapped to the user identification information based on the identification information on the first speech signal not being related to at least one piece of the identification information mapped to the user identification information.

[0339] Additionally, according to one aspect of the present disclosure, the controller **470** may map the identification information on the first speech signal to the user identification information and store the mapped information in the database **490**.

[0340] Additionally, according to one aspect of the present disclosure, the controller **470** may transmit, to the electronic device **100**, a result of determining that a current user is the same as a previous user based on the identification information on the first speech signal being related to at least one piece of the identification information mapped to the user identification information, and transmit, to the electronic device **100**, a result of determining that the current user is different from the previous user based on the identification information on the first speech signal not being related to at least one piece of the identification information mapped to the user identification information.

[0341] Additionally, according to one aspect of the present disclosure, the controller **470** may map the speech signal to the user identification information and store the mapped signal in the database **490** based on the converted text and the preset text being related to each other.

[0342] Additionally, according to one aspect of the present disclosure, the controller **470** may generate identification information related to the speech signal mapped to the user identification information using a first algorithm, and change the identification information mapped to the user identification information and generated using a second algorithm to the identification information generated using the first algorithm.

[0343] The system **10** according to one aspect of the present disclosure includes the electronic device **100** and the server **400**, wherein the electronic device **100** is configured to display a preset text related to registration of identification information while performing a process of registering the identification information related to speech for a user account logged in to the server **100**, upon receiving a speech signal related to the preset text, transmit data including the speech signal to the server **400**, complete the process of registering the identification information based on processing of the speech signal related to the preset text by the server **400**, and stop the process of registering the identification information upon receiving a predetermined input related to speech recognition from the remote control device **200** during the process of registering the identification information, and the server **400** is configured to convert the speech signal included in the data received from the electronic device **100** into text, generate identification information related to the speech signal based on the converted text and the preset text being related to each other, map the generated identification information to user identification information related to the user account and store the mapped information in a database **490**, and maintain or delete the identification information mapped to the user identification information upon receiving a notification regarding interruption of the process of registering the identification information from the electronic device **100** during the process of registering the identification information.

[0344] The attached drawings are only for easy understanding of the embodiments disclosed in this specification, and the technical idea disclosed in this specification is not limited by the attached drawings, and all changes, equivalents, and changes included in the technical scope of the present disclosure are not limited thereby.

[0345] Meanwhile, the operating method of the present disclosure may be implemented as processor-readable code on a processor-readable recording medium. Processor-readable recording media include all types of recording devices that store data that may be read by a processor. Examples of processor-readable recording media include a ROM, a RAM, a CD-ROM, a magnetic tape, a floppy disk, and an optical data storage device, and also include those implemented in the form of a carrier wave, such as transmission through the Internet. Additionally, a processor-readable recording medium is distributed in a computer system connected to a network, and thus processor-readable code may be stored and executed in a distributed manner.

[0346] The electronic device, the server, and the system including the same according to the present disclosure have the following effects.

[0347] According to at least one embodiment of the present disclosure, it is possible to register identification information on user voice in a user account.

[0348] Additionally, according to at least one embodiment of the present disclosure, it is possible to identify a user based on user voice.

[0349] Additionally, according to at least one embodiment of the present disclosure, it is possible to log in to an account of a user identified based on user voice.

[0350] Additionally, according to at least one embodiment of the present disclosure, it is possible to maintain the continuity of identification information on a user who have previously attempted registration in a process of registering identification information on user voice to a user account.

[0351] Although the preferred embodiments of the present disclosure have been disclosed for illustrative purposes, those skilled in the art will appreciate that various modifications, additions and substitutions are possible, without departing from the scope of the disclosure as disclosed in the accompanying claims.

What is claimed is:

1. An electronic device comprising:

a display;

an external device interface configured to communicate with a remote control device;

a network interface configured to communicate with a server;

a user input interface configured to transmit a signal related to a user input; and

a controller,

wherein the controller is configured to:

display a preset text related to registration of identification information through the display while performing a process of registering the identification information related to speech for a user account logged in to the server;

upon receiving a speech signal related to the preset text through the user input interface, transmit data including the speech signal to the server;

complete the process of registering the identification information based on processing of the speech signal related to the preset text by the server; and

interrupt the process of registering the identification information upon receiving a predetermined input related to speech recognition from the remote control device during the process of registering the identification information.

2. The electronic device of claim **1**, wherein the identification information includes a feature vector with respect to a voiceprint of the speech.

3. The electronic device of claim **1**, wherein the controller is configured to:

transmit, to the server, data including a speech signal related to the predetermined input and received from the remote control device after receiving the predetermined input, and

perform an operation related to the predetermined input based on processing of the speech signal related to the predetermined input by the server.

4. The electronic device of claim **1**, wherein the controller is configured to:

determine whether the identification information has been temporarily stored in the server upon starting the process of registering the identification information for the user account,

display a predetermined number of pieces of text in stages based on the identification information having not been temporarily stored in the server, and

display some of the plurality of pieces of text in stages excluding text related to the identification information temporarily stored in the server based on the identification information having been temporarily stored in the server.

5. The electronic device of claim **1**, wherein the controller is configured to:

display a notification regarding interruption of the process of registering the identification information through the display based on termination of the operation related to the predetermined input, and

determine whether to start the process of registering the identification information based on the user input received through the user input interface.

6. A server comprising:

a communication interface configured to communicate with an electronic device;

a database; and

a controller,

wherein the controller is configured to:

convert a speech signal included in data received from the electronic device into text while performing a process of registering identification information related to speech for a user account logged in to the server;

generate identification information related to the speech signal based on the converted text being related to preset text related to registration of the identification information;

map the generated identification information to user identification information related to the user account and store the mapped information in the database; and

maintain or delete the identification information mapped to the user identification information upon receiving a notification regarding interruption of the process of registering the identification information from the electronic device during the process of registering the identification information.

7. The server of claim 6, wherein the identification information includes a feature vector with respect to a voiceprint of the speech.

8. The server of claim 6, wherein the controller is configured to:

maintain the identification information mapped to the user identification information based on a number of pieces of identification information mapped to the user identification information being equal to or greater than a preset minimum number, and

delete the identification information mapped to the user identification information based on the number of pieces of identification information mapped to the user identification information being less than the preset minimum number.

9. The server of claim 6, wherein the controller is configured to:

determine whether the identification information is mapped to the user identification information based on the process of registering the identification information being started for the user account,

transmit a predetermined number of pieces of text to the electronic device in stages based on the identification information not being mapped to the user identification information, and

transmit some of the plurality of pieces of text excluding text related to the identification information mapped to the user identification information to the electronic device in stages based on the identification information being mapped to the user identification information.

10. The server of claim 6, wherein the controller is configured to:

determine whether the identification information is mapped to the user identification information based on the process of registering the identification information being started for the user account,

transmit first text excluding the text related to the identification information mapped to the user identification information among the plurality of pieces of text to the electronic device based on the identification information being mapped to the user identification information,

generate identification information on a first speech signal related to the first text and included in data received from the electronic device,

maintain the identification information mapped to the user identification information based on the identification information on the first speech signal being related to at least one piece of the identification information mapped to the user identification information, and

delete the identification information mapped to the user identification information based on the identification information on the first speech signal not being related to at least one piece of the identification information mapped to the user identification information.

11. The server of claim 10, wherein the controller is configured to map the identification information on the first speech signal to the user identification information and store the mapped information in the database.

12. The server of claim 10, wherein the controller is configured to:

transmit, to the electronic device, a result of determining that a current user is the same as a previous user based on the identification information on the first speech signal being related to at least one piece of the identification information mapped to the user identification information, and

transmit, to the electronic device, a result of determining that the current user is different from the previous user based on the identification information on the first speech signal not being related to at least one piece of the identification information mapped to the user identification information.

13. The server of claim 6, wherein the controller is configured to map the speech signal to the user identification information and stores the mapped signal in the database based on the converted text and the preset text being related to each other.

14. The server of claim 13, wherein the controller is configured to:

generate identification information related to the speech signal mapped to the user identification information using a first algorithm, and

change the identification information mapped to the user identification information and generated using a second algorithm to the identification information generated using the first algorithm.

15. A system comprising:

an electronic device; and

a server,

wherein the electronic device is configured to:

display a preset text related to registration of identification information while performing a process of registering the identification information related to speech for a user account logged in to the server;

upon receiving a speech signal related to the preset text, transmit data including the speech signal to the server;

complete the process of registering the identification information based on processing of the speech signal related to the preset text by the server; and

stop the process of registering the identification information upon receiving a predetermined input related to speech recognition from a remote control device during the process of registering the identification information, and

the server is configured to:

convert the speech signal included in the data received from the electronic device into text;

generate identification information related to the speech signal based on the converted text and the preset text being related to each other;

map the generated identification information to user identification information related to the user account and store the mapped information in a database; and

maintain or delete the identification information mapped to the user identification information upon receiving a notification regarding interruption of the process of registering the identification information from the electronic device during the process of registering the identification information.

* * * * *