United States Patent Application Publication	20250260686
Kind Code	A1
Publication Date	August 14, 2025
Inventor(s)	Lee; Sangseok et al.

ELECTRONIC DEVICE AND SYSTEM INCLUDING SAME

Abstract

An electronic device including a display; a network interface configured to communicate with a server; a user input interface configured to receive speech uttered by a user; and a controller configured to determine whether the speech uttered by the user is related to a user account logged in to the server, in response to determining the speech uttered by the user is unrelated to the user account logged in to the server, display a first response on the display including an indicator indicating the user who has uttered the speech is not the user logged in to the user account, and in response to determining the speech uttered by the user is related to the user account logged in to the server, display a second response on the display indicating the user who uttered the speech is currently logged in to the server.

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

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

Family ID: **1000008264121**

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

Appl. No.: **18/931599**

Filed: **October 30, 2024**

Foreign Application Priority Data

KR	10-2024-0019974	Feb. 08, 2024
KR	10-2024-0077175	Jun. 13, 2024

Publication Classification

Int. Cl.: **H04L9/40** (20220101); **G10L17/06** (20130101); **G10L17/22** (20130101)

U.S. Cl.:

CPC **H04L63/0861** (20130101); **G10L17/06** (20130101); **G10L17/22** (20130101);

Background/Summary

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to Korean Patent Application Nos. 10-2024-0019974 and 10-2024-0077175, filed in the Republic of Korea on Feb. 8, 2024, and Jun. 13, 2024, respectively, the entire contents of which is hereby expressly incorporated by reference into the present application.

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0002] The present disclosure relates to an electronic device and a system including the same, and more specifically, to an electronic device that utilizes voice recognition technology and a system including the same.

2. Description of the Related Art

[0003] With the recent development of technology, research on voice recognition technology for processing speech is being actively conducted. In particular, research on voice 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.

[0004] Voice 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 can directly recognize and process user's speech and operate according to the command corresponding to the speech, or can send the speech to a server that processes speech and then operate according to a command corresponding to the speech received from the server.

[0005] 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 instance, service providers use user information managed for each account to provide optimal functions or information tailored to the user.

[0006] 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 can 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

[0007] 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.

[0008] It is another object of the present disclosure to provide an electronic device capable of registering identification information on a user voice in a user account and a system including the same.

[0009] It is a further object of the present disclosure to provide an electronic device capable of identifying a user based on a user voice and a system including the same.

[0010] It is a further object of the present disclosure to provide an electronic device capable of allowing a user to log in to a user account identified based on user voice and a system including the

same.
[0011] It is a further object of the present disclosure to provide an electronic device capable of accurately notifying a user of switching of a logged-in user account and a system including the same.
[0012] In accordance with the present disclosure, the above and other objects can be accomplished by the provision of an electronic device including a display, a network interface configured to communicate with a server, a user input interface configured to transmit a speech signal corresponding to speech uttered by a user, and a controller.
[0013] The controller determines whether the user who has uttered the speech is related to a user account logged in to the server, displays a first response including an indicator indicating the user who has uttered the speech through the display based on the user who has uttered the speech being unrelated to the user account logged in to the server, and displays a second response not including the indicator through the display based on the user who has uttered the speech being related to the user account logged in to the server.
[0014] 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 transmit data including a speech signal related to speech uttered by a user to the server, determine whether the user who has uttered the speech is related to a user account logged in to the server based on a result of processing the speech signal received from the server, display a first response including an indicator indicating the user who has uttered the speech through a display based on the user who has uttered the speech being unrelated to the user account logged in to the server, and display a second response not including the indicator through the display based on the user who has uttered the speech being related to the user account logged into the server, and the server is configured to generate identification information for the speech signal included in the data received from the electronic device, determine first identification information related to the identification information for the speech signal among identification information mapped to user identification information related to user accounts stored in a database of the server, and transmit the result of processing the speech signal including first user identification information mapped to the first identification information to the electronic device.

Description

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. **13** are diagrams referenced in description of a process of registering identification information on a user voice in a user account according to an embodiment of the present disclosure;
[0025] FIGS. **14** and **15** are flowcharts of a method of operating an electronic device according to an embodiment of the present disclosure; and
[0026] FIGS. **16** to **29** are diagrams referenced in description of responses to user voice 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” can 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 combin ations 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. Referring to FIG. **1**, the system **10** can include an electronic device **100** and/or a server **400**.
[0032] The electronic device **100** can transmit/receive data to/from at least one server **400**. For example, the electronic device **100** can transmit/receive data to/from the at least one server **400** via a network **300** such as the Internet.
[0033] According to an embodiment, the at least one server **400** can 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.
[0034] The electronic device **100** can 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.
[0035] The image display device **100***a* can 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 corresponding to video signals, such as a TV, a laptop computer, or a monitor.
[0036] The image display device **100***a* can 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* can correspond to a broadcast reception device.
[0037] The image display device **100***a* can receive broadcast signals wirelessly through an antenna, or can receive broadcast signals through a cable. For example, the image display device **100***a* can receive terrestrial broadcast signals, satellite broadcast signals, cable broadcast signals, and Internet protocol Television (IPTV) broadcast signals.
[0038] FIG. **2** is an internal block diagram of the electronic device of FIG. **1**. Referring to FIG. **2**, the electronic device **100** can 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**. The broadcast receiver **105** can include a tuner **110** and a demodulator **120**.
[0039] Meanwhile, the electronic device **100** can 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**.
[0040] The tuner **110** can select a broadcast signal corresponding to a channel selected by a user or broadcast signals of all previously stored channels among broadcast signals received through an antenna or a cable. The tuner **110** can convert the selected broadcast signals into intermediate frequency signals or baseband video or audio signals.
[0041] For example, if a selected broadcast signal is a digital broadcast signal, the tuner **110** can 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** can process digital broadcast signals or analog broadcast signals. The analog base band video or audio signal (CVBS/SIF) output from the tuner **110** can be directly input to the controller **170**. Meanwhile, the tuner **110** can 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. 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 can also be adopted.
[0042] The demodulator **120** can receive a digital IF signal (DIF) converted by the tuner **110** and perform a demodulation operation. The demodulator **120** can output a stream signal TS after performing demodulation and channel decoding. Here, the stream signal can be a multiplexed video signal, audio signal, or data signal.
[0043] The stream signal output from the demodulator **120** can be input to the controller **170**. After performing demultiplexing and video/audio signal processing, the controller **170** can output video through the display **180** and output audio through the audio output part **185**.
[0044] The external device interface **130** can transmit/receive data to/from a connected external device. To this end, the external device interface **130** can include an A/V input/output part. The external device interface **130** can 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 can also perform input/output operations with respect to external devices.
[0045] In addition, the external device interface **130** can 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**.
[0046] The A/V input/output part can receive video and audio signals from an external device. For example, the A/V input/output part can 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 can be transmitted to the controller **170**. Here, analog signals input through the CVBS terminal and the S-video terminal can be converted into digital signals through an analog-to-digital converter and transmitted to the controller **170**.
[0047] The external device interface **130** can include a wireless communication part for short-distance wireless communication with other electronic devices. The external device interface **130**

can exchange data with a neighboring mobile terminal through the wireless communication part. For example, the external device interface **130** can receive device information, executing application information, application images, and the like from the mobile terminal in a mirroring mode. The external device interface **130** can perform short-range wireless communication using Bluetooth, radio frequency identification (RFID), infrared data association (IrDA), ultra-wideband (UWB), ZigBee, and the like.

[0048] The network interface **135** can provide an interface for connecting the electronic device **100** to a wired/wireless network including the Internet. The network interface **135** can include a communication module for connection to a wired/wireless network. For example, the network interface **135** can 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).

[0049] The network interface **135** can transmit/receive data to/from other users or other electronic devices through a connected network or another network linked to the connected network. The network interface **135** can receive web content or data provided by content providers or network operators. That is, the network interface **135** can receive content such as movies, advertisements, games, VOD, and broadcasting and information related thereto provided from content providers or network providers through networks.

[0050] The network interface **135** can receive firmware update information and update files provided by network operators, and can transmit data to the Internet, content providers, or network operators. The network interface **135** can select and receive a desired application from among applications open to the public through a network.

[0051] The storage **140** can store programs for processing and controlling each signal in the controller **170** and can store processed video, audio, or data signals. For example, the storage **140** can store application programs designed for the purpose of performing various tasks that can be processed by the controller **170** and selectively provide some of the stored application programs at the request of the controller **170**.

[0052] Programs stored in the storage **140** are not particularly limited as long as they can be executed by the controller **170**. The storage **140** can execute a function of temporarily storing video, voice, or data signals received from an external device through the external device interface **130**. The storage **140** can store information on a predetermined broadcast channel through a channel memory function such as a channel map.

[0053] 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** can be included in the controller **170**. The storage **140** can 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” can be used interchangeably.

[0054] The user input interface **150** can transmit a signal input by a user to the controller **170** or transmit a signal from the controller **170** to the user. For example, the user input interface **150** can 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 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 that senses a user's gesture to the controller **170**, or transmit signals from the controller **170** to the sensor.

[0055] The input part **160** can be provided on one side of the main body of the electronic device **100**. For example, the input part **160** can include a touch pad, physical buttons, and the like. The input part **160** can receive various user commands related to the operation of the electronic device **100** and transmit control signals corresponding to the input commands to the controller **170**. The input part **160** can include at least one microphone and can receive a user voice through the

microphone.
[0056] The controller **170** can include at least one processor and can control the overall operation of the electronic device **100** using the processor included therein. Here, the processor can be a general processor such as a central processing unit (CPU). The processor can be a dedicated device such as an ASIC or another hardware-based processor.
[0057] The controller **170** can 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. The display **180** can 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.
[0058] The display **180** can include a display panel having a plurality of pixels. The plurality of pixels provided in the display panel can include RGB subpixels. Alternatively, the plurality of pixels provided in the display panel can include RGBW subpixels. The display **180** can 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.
[0059] The display **180** can be a plasma display panel (PDP), a liquid crystal display (LCD), an organic light emitting diode (OLED) display, or a flexible display, and can also be a 3D display. 3D displays **180** can be classified into a glasses-free type and a glasses type. Meanwhile, the display **180** can be configured as a touch screen and used as an input device in addition to an output device.
[0060] The audio output part **185** receives the audio signal processed by the controller **170** and outputs the same as audio. A video signal processed by the controller **170** can be input to the display **180** and displayed as an image corresponding to the video signal. Additionally, the video signal processed by the controller **170** can be input to an external output device through the external device interface **130**.
[0061] An audio signal processed by the controller **170** can be output as sound to the audio output part **185**. Additionally, the audio signal processed by the controller **170** can be input to an external output device through the external device interface **130**.
[0062] Further, the controller **170** can include a demultiplexer, an image processor, etc. In addition, the controller **170** can control overall operations of the electronic device **100**. For example, the controller **170** can control the tuner **110** to select (tune to) a broadcast corresponding to a channel selected by the user or a previously stored channel.
[0063] Additionally, the controller **170** can control the electronic device **100** using a user command input through the user input interface **150** or an internal program. Meanwhile, the controller **170** can control the display **180** to display an image. Here, the image displayed on the display **180** can be a still image or a video, and can be a 2D image or a 3D image.
[0064] Further, the controller **170** can cause a predetermined 2D object to be displayed in an image displayed on the display **180**. For example, the object can 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.
[0065] Meanwhile, the electronic device **100** can further include an imaging device. The imaging device can capture an image of the user. The imaging device can be implemented as a single camera, but the present disclosure is not limited thereto and the imaging device can also be implemented as a plurality of cameras. Meanwhile, the imaging device can be embedded in the electronic device **100** at the top of the display **180** or can be disposed separately. Image information captured by the imaging device can be input to the controller **170**.
[0066] The controller **170** can recognize a location of the user based on images captured by the imaging device. For example, the controller **170** can ascertain the distance (z-axis coordinate) between the user and the electronic device **100**. In addition, the controller **170** can ascertain the x-axis coordinate and y-axis coordinate in the display **180** corresponding to the location of the user.

The controller **170** can detect a user's gesture based on images captured by the imaging device, each signal detected by a sensor, or a combination thereof.
[0067] The power supply **190** can supply corresponding power throughout the electronic device **100**. In particular, the power supply **190** can supply power to the controller **170**, which can 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. Specifically, the power supply **190** can include a converter that converts AC power to DC power and a DC/DC converter that converts a DC power level.
[0068] The remote control device **200** can transmit user input to the user input interface **150**. To this end, the remote control device **200** can use Bluetooth, radio frequency (RF) communication, infrared communication, ultra-wideband (UWB), ZigBee, and the like. Additionally, the remote control device **200** can 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**.
[0069] The electronic device **100** described above can be a stationary or mobile digital broadcast receiver capable of receiving digital broadcasting. 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 can be integrated, added, or omitted according to the specifications of the electronic device **100** that is actually implemented.
[0070] That is, two or more components can be combined into one component, or one component can 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.
[0071] Next, FIG. **3** is a diagram referenced in description of the server of FIG. **1**. Referring to FIG. **3**, the server **400** can 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** can be configured as one server.
[0072] The relay server **410** can communicate with the electronic device **100**. The relay server **410** can 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** can 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**.
[0073] The STT server **420** can receive speech data. The STT server **420** can convert the speech data into text data. The STT server **420** can transmit the text data to the electronic device **100** via the relay server **410**. The STT server **420** can be called an automatic speech recognition (ASR) server.
[0074] The STT server **420** can increase the accuracy of speech-to-text conversion using a language model. A language model can refer to a model that can calculate the probability of a sentence or the probability of the next word appearing when previous words are provided. For example, the language model can include probabilistic language models such as Unigram model, Bigram model, and N-gram model. That is, the STT server **420** can determine whether text data has been appropriately converted from speech data, and accordingly, increase the accuracy of conversion to text data.
[0075] The NLP server **430** can receive text data. The NLP server **430** can perform intent analysis on the text data based on the received text data. The NLP server **430** can transmit intent analysis information indicating the result of intent analysis to the electronic device **100** via the relay server **410**.
[0076] According to an embodiment, the NLP server **430** can 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 corresponding 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 can 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.

[0077] The user identification server **440** can receive voice data. The user identification server **440** can extract voice features based on the voice data. Here, the voice features can 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 FIGS. **4** and **5**.

[0078] The user identification server **440** can obtain a voice feature vector from the voice features. The user identification server **440** can 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.

[0079] The user identification server **440** can determine a similarity between a plurality of feature vectors. The user identification server **440** can 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 corresponding to first text and a second vector corresponding to second text can be created. A cosine similarity between the first vector and the second vector can be calculated based on Formula 1 below.

[00001] $$\cos(\Theta) = \frac{A \cdot B}{\|A\| \, \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$ [Formula1]

[0080] 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 can be calculated by dividing the dot product of two vectors by the product of the magnitudes of the vectors. Cosine similarity can range from −1 to 1, and two vectors are determined to be similar as the cosine similarity therebetween is closer to 1.

[0081] The user identification server **440** can 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 corresponding to the first voice input and a second feature vector corresponding to the second voice input is equal to or greater than a predetermined standard, the user identification server **440** can determine that the user who has uttered the first voice input and the user who has uttered the second voice input are the same.

[0082] According to an embodiment, the user identification server **440** can 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** can 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.

[0083] The user identification server **440** can store voice data. The user identification server **440** can store data on voiceprint (hereinafter, voiceprint information). Here, voiceprint information can include a voice feature vector and/or a vector obtained by processing the voice feature vector.

[0084] The user identification server **440** can store a voice database. The voice database regarding voices can include unique identification information corresponding to the electronic device **100** (hereinafter referred to as device identification information), unique identification information corresponding 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.
[0085] The device identification information, user identification information, voice data, and voiceprint information included in the voice database can 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 voice data, and/or a plurality of pieces of voiceprint information can be mapped to user identification information. That is, it can be interpreted that device identification information, voice 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 voice data and a plurality of pieces of voiceprint information are all mapped to user identification information included in the voice database will be described.
[0086] The user identification server **440** can update voiceprint information included in a voice database based on voice data included in the voice database. For example, the user identification server **440** can generate voiceprint information corresponding to voice data included in the voice database using an algorithm different from a previously used algorithm. Here, the user identification server **440** can change the voiceprint information included in the voice database to the newly generated voiceprint information.
[0087] The account server **450** can manage data regarding user accounts. The account server **450** can 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. The account server **450** can store a database regarding user accounts. The database regarding user accounts can 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.
[0088] The account server **450** can communicate with the electronic device **100**. For example, the account server **450** can create and register a user account based on data from the electronic device **100**. For example, the account server **450** can approve login of a user account based on an ID and a password received from the electronic device **100**.
[0089] FIG. **4** is a block diagram for describing the configuration of the server according to an embodiment of the present disclosure. Referring to FIG. **4**, the server **400** can include a preprocessor **460**, a controller **470**, a communication part **480**, and/or a database **490**.
[0090] The preprocessor **460** can preprocess speech received through the communication part **480** or speech stored in the database **490**. The preprocessor **460** can be implemented as a separate chip from the controller **470** or can be implemented as a chip included in the controller **470**. The preprocessor **460** can 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.
[0091] If the preprocessor **460** is provided in the electronic device **100**, the preprocessor **460** can recognize a startup word for activating voice recognition of the electronic device **100**. The preprocessor **460** can convert the startup word received through the user input interface **150** into text data, and if the converted text data is text data corresponding to a pre-stored startup word, determine that the startup word is recognized. The preprocessor **460** can convert the noise-removed voice signal into a power spectrum.
[0092] A power spectrum can 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. 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**.
[0093] In particular, 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. FIG. **5** shows a voice signal **510**. The voice signal **510** can be a signal received from an external device or can be a signal previously stored in the memory **170**. The x-axis of the voice signal **510** represents time, and the y-axis represents amplitude.
[0094] A power spectrum processor **463** can 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** can 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.
[0095] Referring back to FIG. **4**, the functions of the preprocessor **460** and the controller **470** described in FIG. **4** can also be performed in the NLP server **430**. The preprocessor **460** can include a wave processor **461**, a frequency processor **462**, the power spectrum processor **463**, a speech-to-text (STT) converter **464**, and the like.
[0096] The wave processor **461** can extract the waveform of speech. The frequency processor **462** can extract the frequency band of the speech. The power spectrum processor **463** can extract the power spectrum of the speech. A power spectrum can 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.
[0097] The STT converter **464** can convert speech into text. The STT converter **464** can convert speech in a specific language into text in that language. The controller **470** can control the overall operation of the server **400**. The controller **470** can include a speech analyzer **471**, a text analyzer **472**, a feature clustering part **473**, a text mapper **474**, and/or a speech synthesizer **475**.
[0098] The speech analyzer **471** can 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 can 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 can further include the timbre of the speaker.
[0099] The text analyzer **472** can 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** can extract the phrase with a different tone as a main expression phrase. The text analyzer **472** can determine that the tone has changed when the frequency band between phrases has changed more than a preset band. The text analyzer **472** can extract key words from phrases in the converted text. A key word can be a noun present in a phrase, but this is merely an example.
[0100] The feature clustering part **473** can classify the speech type of the speaker using the speech characteristic information extracted by the speech analyzer **471**. The feature clustering part **473** can 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** can classify the speech type of the speaker using an attention technique of a deep learning model.
[0101] The text mapper **474** can translate text converted into a first language into text in a second language. The text mapper **474** can map the text translated into the second language with the text in the first language. The text mapper **474** can map main expressions constituting the text in the first language to corresponding phrases in the second language. The text mapper **474** can map a speech type corresponding 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.
[0102] The speech synthesizer **475** can 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. The controller **470** can determine the speech characteristics of the user using one or more of the transmitted text data or the power spectrum **520**. Speech characteristics of a user can include the sex, pitch, tone, speech topic, speech rate, and voice volume of the user.

[0103] The controller **470** can obtain the frequency of the voice signal **510** and the amplitude corresponding to the frequency. The controller **470** can 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** can determine that the user is male.

[0104] If the frequency band of the power spectrum **520** is within a preset second frequency band range, the controller **470** can determine that the user is female. Here, the second frequency band range can be higher than the first frequency band range. The controller **470** can determine the pitch of voice using the frequency band of the power spectrum **520**. For example, the controller **470** can determine the pitch of the voice based on the amplitude within a specific frequency band.

[0105] The controller **470** can determine the user's tone using the frequency band of the power spectrum **520**. For example, the controller **470** can 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.

[0106] The controller **470** can determine the user's speech rate based on the number of syllables uttered per unit time from the converted text data. The controller **470** can determine the topic of the user's speech using the Bag-Of-Word Model technique for the converted text data.

[0107] 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** can classify the topic of the user's speech as exercise.

[0108] The controller **470** can determine the topic of the user's speech from the text data using a known text categorization technique. The controller **470** can extract keywords from the text data and determine the topic of the user's speech. The controller **470** can determine the user's voice volume by considering amplitude information in the entire frequency band. For example, the controller **470** can determine the user's voice volume based on the average or weighted average of amplitudes in each frequency band of the power spectrum **470**.

[0109] The communication part **480** can communicate with an external server by wire or wirelessly. The communication part **480** can communicate with the electronic device **100** by wire or wirelessly. The database **490** can store speech in first language included in content. The database **490** can store synthetic speech in which speech in the first language has been converted into speech in the second language. The database **490** can store first text corresponding to speech in the first language and second text in which the first text has been translated into the second language. The database **490** can store various learning models required for voice recognition.

[0110] Meanwhile, the controller **170** of the electronic device **100** illustrated in FIG. **2** can include the preprocessor **460** and the controller **470** illustrated in FIG. **4**. That is, the controller **170** of the electronic device **100** can perform the functions of the preprocessor **460** and the controller **470**.

[0111] Next, FIG. **6** is a block diagram illustrating a configuration of a controller for voice recognition and synthesis of an image display device according to an embodiment of the present disclosure. That is, the voice recognition and synthesis process illustrated in FIG. **6** can be performed by the controller **170** of the electronic device **100** without using the server.

[0112] Referring to FIG. **6**, the processor **170** of the electronic device **100** can include an STT engine **610**, an NLP engine **620**, and a speech synthesis engine **630**. Each engine can be either hardware or software. The STT engine **610** can perform the function of the STT server **420** of FIG.

5. That is, the STT engine **610** can convert voice data into text data.
[0113] The NLP engine **620** can perform the function of the NLP server **430** shown in FIG. **5**. That is, the NLP engine **620** can obtain intent analysis information indicating the speaker's intention from the converted text data.
[0114] The speech synthesis engine **630** can perform a function of a speech synthesis server. The speech synthesis engine **630** can search a database for syllables or words corresponding to given text data and synthesize a combination of the searched syllables or words to generate synthetic speech. The speech synthesis engine **630** can include a preprocessing engine **631** and a TTS engine **632**.
[0115] The preprocessing engine **631** can 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** can perform a cleansing operation to remove unnecessary characters and symbols to eliminate noise. Thereafter, the preprocessing engine **631** can generate the same word token by integrating word tokens with different expression methods. Thereafter, the preprocessing engine **631** can remove meaningless word tokens (stopwords). The TTS engine **632** can synthesize speech corresponding to the preprocessed text data and generate synthetic speech.
[0116] Next, FIG. **7** is a flowchart of a method of operating an electronic device according to an embodiment of the present disclosure. Referring to FIG. **7**, the electronic device **100** can determine whether a user account is logged in to the server **400** in operation S**701**. For example, a user can log in to the server **400** with a user account by entering the user account ID and password.
[0117] 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** can include user identification information corresponding to the user account in a user list. For example, when three different user accounts log in to the server **400** using the electronic device **100**, the user list stored in the electronic device **100** can include three different pieces of user identification information.
[0118] In operation S**702**, the electronic device **100** can 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 can include voiceprint information stored in the user identification server **440**. For example, the server **400** can 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**.
[0119] According to an embodiment, the server **400** can 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 corresponding 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 can be 0.
[0120] According to an embodiment, the server **400** can 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, for a user account for which a voice ID has been registered, six different pieces of voiceprint information can be mapped to user identification information. For example, for a user account for which a voice ID has not been registered, five or fewer voiceprint information can be mapped to user identification information.
[0121] According to an embodiment, a flag value indicating whether a voice ID has been registered can be mapped to user identification information stored in the server **400**. Here, user identification information to which a flag value is mapped can be stored in the user identification server **440** and/or the account server **450**. The server **400** can 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 can be 0 for a user account for which a voice ID has not been registered, and a flag value mapped to user identification information can be 1 for a user account for which a voice ID has been registered.
[0122] When the voice ID has not been registered with respect to the user account, the electronic device **100** can 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** can 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**.
[0123] The electronic device **100** can output preset text in operation S**704**. The electronic device **100** can 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** can output preset text through the display **180**.
[0124] According to an embodiment, the server **400** can 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** can output the preset text received from the server **400**.
[0125] The electronic device **100** can determine whether speech with respect to the preset text is input in operation S**705**. For example, the electronic device **100** can determine whether speech is input through a microphone included in the input part **160** within a preset time. Here, the voice signal corresponding to the speech input through the microphone can be transmitted to the controller **170** through the user input interface **150**. For example, the electronic device **100** can determine whether data containing a voice signal corresponding to speech uttered by the user is received from the remote control device **200** within a preset time.
[0126] When speech with respect to the preset text is input, the electronic device **100** can transmit voice data including the voice signal corresponding to the speech to the server **400** in operation S**706**. Here, the electronic device **100** can 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 voice data.
[0127] The server **400** can convert the voice signal included in the voice data received from the electronic device **100** into text. The server **400** can determine whether the text converted from the voice signal and the preset text correspond to each other. For example, the server **400** can determine whether the text converted from the voice signal and the preset text correspond to each other based on the similarity therebetween.
[0128] The server **400** can generate voiceprint information corresponding to the voice signal when the text converted from the voice signal and the preset text correspond to each other. The server **400** can map the voiceprint information generated with respect to the preset text to the user identification information and store the same. The server **400** can map the voice data received with respect to the preset text to the user identification information and store the same.
[0129] The electronic device **100** can 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** can notify the electronic device **100** of success of speech processing. For example, when the voiceprint information corresponding to the voice signal has been generated, the server **400** can notify the electronic device **100** of success of speech processing.
[0130] Meanwhile, in operation S**708**, the electronic device **100** can determine whether the user re-attempts 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** can re-attempt to input speech based on a user input from the user re-attempting to input speech. Here, the electronic device **100** can output the preset text again.
[0131] In operation S**709**, if speech processing for the preset text is successful, the electronic

device **100** can 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 can be completed. Meanwhile, when processing for five pieces of preset text is completed, the electronic device **100** can output the last preset text.

[0132] The electronic device **100** can 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** can output a screen indicating completion of voice ID registration through the display **180**. For example, the electronic device **100** can transmit data indicating completion of voice ID registration to the account server **450**.

[0133] Next, FIG. **8** is a flowchart of a method of operating a system according to an embodiment of the present disclosure. Referring to FIG. **8**, the electronic device **100** can log in to the server **400** using a user account in operation S**801**.

[0134] The electronic device **100** can start a process of registering a voice ID in operation S**802**, output first text among a plurality of pieces of preset text in operation S**803** and receive first speech for the first text in operation S**804**. The electronic device **100** can then transmit first speech data including a speech signal corresponding to the first speech to the server **400** in operation S**805**.

[0135] The server **400** can process the first speech for the first text based on the first speech data received from the electronic device **100** in operation S**806**. The server **400** can convert the speech signal corresponding to the first speech included in the first speech data received from the electronic device **100** into text. The server **400** can determine whether the text converted from the speech signal corresponding to the first speech and the first text correspond to each other.

[0136] In addition, the server **400** can notify the electronic device **100** of completion of processing for the first speech in operation S**807**. For example, the server **400** can 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 corresponding to the first speech corresponds to the first text.

[0137] Further, the server **400** can generate first voiceprint information with respect to the first speech based on the speech signal corresponding to the first speech based on the fact that the text converted from the speech signal corresponding to the first speech corresponds to the first text. The server **400** can store the first speech data and the first voiceprint information with respect to the first speech in operation S**808**. The server **400** can map the first speech data and first voiceprint information to the user identification information corresponding to the logged-in user account and store the same.

[0138] The electronic device **100** can output the second to fifth pieces of text in stages. The electronic device **100** can sequentially receive second to fifth speeches corresponding to the second to fifth pieces of text. The electronic device **100** can sequentially transmit second to fifth pieces of speech data corresponding to the second to fifth speeches to the server **400**.

[0139] The server **400** can process the second to fifth speeches based on the second to fifth pieces of speech data received from the electronic device **100**. Additionally, the server **400** can sequentially generate and store second to fifth pieces of speech information corresponding to the second to fifth speeches.

[0140] The electronic device **100** can output sixth text from among a plurality of pieces of preset text in operation S**809**, receive sixth speech with respect to the sixth text in operation S**810**, and transmit sixth speech data including a speech signal corresponding to the sixth speech to the server **400** in operation S**811**.

[0141] The server **400** can process the sixth speech with respect to the sixth text based on the sixth speech data received from the electronic device **100** in operation S**812**. The server **400** can convert a speech signal corresponding to the sixth speech included in the sixth speech data received from the electronic device **100** into text. The server **400** can determine whether the text converted from the speech signal corresponding to the sixth speech and the sixth text correspond to each other. The server **400** can notify the electronic device **100** of completion of processing for the sixth speech in

operation S**813**.
[0142] Meanwhile, when the text converted from the speech signal corresponding to the sixth speech and the sixth text correspond to each other, the server **400** can generate sixth voiceprint information regarding the sixth speech based on the speech signal corresponding to the sixth speech.
[0143] The server **400** can store the sixth speech data and the sixth voiceprint information regarding the sixth speech in operation S**814**. The server **400** can map the sixth speech data and the sixth voiceprint information to the user identification information corresponding to the logged-in user account and store the same. Here, six different pieces of speech data and a plurality of pieces of voiceprint information can be mapped to the user identification information corresponding to the logged-in user account.
[0144] The electronic device **100** can end the process of registering the voice ID in operation S**815**. For example, the electronic device **100** can end the process of registering the voice ID based on completion of processing for the six different pieces of preset text.
[0145] Referring to FIG. **9**, if the user account is not logged in to the server **400**, the electronic device **100** can output a login screen **900** related to logging in to the server **400** through the display **180**. The login screen **900** can 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** corresponding to the remote control device **200**, the electronic device **100** can output a screen for entering an ID and a password. Here, the user can log in to the server **400** with the user account by entering the ID and the password of the user account.
[0146] 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** can output a first account screen **1000** corresponding to the user account for which the voice ID has not been registered. The first account screen **1000** can 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** can start the process of registering a voice ID.
[0147] Referring to FIG. **11**, when a voice ID has been registered in a user account logged in to the server **400**, the electronic device **100** can display a second account screen **1100** corresponding to the user account for which the voice ID has been registered. The second account screen **1100** can 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 can select the activation object **1140** using the pointer **205** to activate or deactivate the use of a function related to voice ID.
[0148] Referring to FIG. **12**, when the object **1020** regarding voice ID registration is selected on the first account screen **1000**, or when the re-registration object **1120** is selected on the second account screen **1100**, the electronic device **100** can 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** can output a text screen for displaying preset text.
[0149] Referring to FIG. **13**, the electronic device **100** can output a text screen **1300** for displaying any one of a plurality of pieces of preset text. The text screen **1300** can 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.
[0150] When the user selects the end object **1310** using the pointer **205**, the process of registering a voice ID can 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 can be deleted. When the user selects the input object **1320** using the pointer **205**, the electronic device **100** can receive speech with respect to text.
[0151] 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** can receive speech with respect to text based on the user input of pressing the predetermined button, received from the remote control device **200**.

[0152] 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** can 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** can correspond to a user input of starting speech recognition for speech received through the remote control device **200**. The electronic device **100** can perform an operation related to speech recognition on speech data including a speech signal received from the remote control device **200**.

[0153] Next, FIGS. **14** and **15** are flowcharts of a method of operating the electronic device according to an embodiment of the present disclosure. Referring to FIG. **14**, the electronic device **100** can connect to the server **400** through the network interface **145** in operation S**1401**.

[0154] The electronic device **100** can receive speech in operation S**1402**. For example, the electronic device **100** can receive speech through a microphone included in the input part **160**. Here, the speech signal corresponding to the speech input through the microphone can be transmitted to the controller **170** through the user input interface **150**. For example, the electronic device **100** can receive data containing a speech signal corresponding to speech uttered by the user from the remote control device **200**.

[0155] When speech is input in operation S**1403**, the electronic device **100** can transmit speech data corresponding to the input speech to the server **400**. Here, the electronic device **100** can transmit the device identification information, a user list, and a language code indicating a type of language to the server **400** along with the speech data.

[0156] The electronic device **100** can receive the result of speech processing from the server **400** in operation S**1404**. For example, the result of speech processing can include text corresponding to the speech, intent analysis information that is the result of performing natural language processing on the speech, and user identification information corresponding to the speech.

[0157] The server **400** can generate voiceprint information with respect to the speech input to the electronic device **100** based on the speech data received from the electronic device **100**. The server **400** can search a database with respect to speech for voiceprint information (hereinafter referred to as candidate voiceprint information) corresponding to user identification information included in the user list received from the electronic device **100**. The server **400** can determine whether the candidate voiceprint information and the generated voiceprint information correspond to each other. The server **400** can determine user identification information to which candidate voiceprint information corresponding to the generated voiceprint information is mapped among candidate voiceprint information as the user identification information corresponding to the speech input to the electronic device **100**. If there is no candidate voiceprint information corresponding to the generated voiceprint information, the server **400** can determine that user identification information corresponding to the speech input to the electronic device **100** is not present.

[0158] The server **400** can transmit the result of processing the speech data received from the electronic device **100** to the electronic device **100**. For example, the server **400** can transmit, to the electronic device **100**, text converted from the speech data received from the electronic device **100**, intent analysis information representing the result of intent analysis performed on the converted text, user identification information corresponding to the speech, and presence or absence of user identification information corresponding to the speech. In operation S**1405**, the electronic device **100** can determine whether the input speech requires user identification based on the result of speech processing.

[0159] According to an embodiment, the electronic device **100** can determine whether the input speech requires user identification based on the type of command corresponding to the speech. For

example, the electronic device **100** can determine that the input speech requires user identification based on the input speech corresponding to a command related to the user account (e.g., user account login, user account switching, or the like). For example, the electronic device **100** can determine that the input speech requires user identification based on the input speech corresponding to a command (e.g., content search, content recommendation, speech recommendation, external device connection, or the like) using user information such as a usage history, a viewing history, and preferred genres. For example, the electronic device **100** can determine that the input speech does not require user identification based on the input speech corresponding to a command for checking general information (e.g., time, weather, etc.) unrelated to the user. For example, the electronic device **100** can determine that the input speech does not require user identification based on the input speech corresponding to a command for adjusting settings (e.g., volume, screen brightness, etc.) of the electronic device **100**.

[0160] According to an embodiment, the electronic device **100** can determine whether the input speech requires user identification based on the type of application corresponding to the speech. For example, the electronic device **100** can determine that the input speech requires user identification based on the input speech corresponding to an application (e.g., an over-the-top media service (OTT service), a social network service (SNS), or the like) linked to the user account. For example, the electronic device **100** can determine that the input speech does not require user identification based on the input speech corresponding to an application (e.g., terrestrial broadcasting, weather, or the like) that is not linked to the user account.

[0161] In operation S**1406**, the electronic device **100** can generate a response related to the user when the input speech requires user identification. For example, when the input speech requires user identification, the electronic device **100** can generate a response including an indicator (hereinafter referred to as a user indicator) indicating the user who has uttered the speech based on user identification information corresponding to the speech included in the speech processing result. Additionally, the electronic device **100** can output a response including a user indicator. Here, the user indicator can include the user's name, nickname, an icon, a photo, and the like corresponding to the user identification information. The user indicator can be stored in the electronic device **100** and/or the server **400**.

[0162] According to an embodiment, when the input speech requires user identification, the electronic device **100** can determine whether to generate a first response including a user indicator or a second response that does not include a user indicator. This will be described with reference to FIG. **15**.

[0163] Referring to FIG. **15**, the electronic device **100** can determine whether the user who has uttered speech is a user preset to use a voice ID based on user identification information corresponding to the speech in operation S**1501**. For example, the electronic device **100** can store information on whether use of a voice ID-related function is activated for each piece of user identification information included in the user list. Here, if the use of the voice ID-related function is activated for the user identification information corresponding to the speech included in the speech processing result, the electronic device **100** can determine that the user who has uttered the speech is a user preset to use the voice ID. If there is no user identification information corresponding to the speech based on the speech processing result, the electronic device **100** can determine that the user who has uttered the speech is not a user preset to use the voice ID.

[0164] If the user who has uttered the speech is a user preset to use the voice ID, the electronic device **100** can determine whether the user who has uttered the speech corresponds to the currently logged-in user account in operation S**1502**. For example, if the user identification information corresponding to the speech corresponds to the user identification information corresponding to the currently logged-in user account, the electronic device **100** can determine that the user who has uttered the speech corresponds to the currently logged-in user account.

[0165] If the user who has uttered the speech does not correspond to the currently logged-in user

account, the electronic device **100** can determine whether switching of the user account is necessary in operation S**1503**. The electronic device **100** can determine whether switching of the user account is necessary based on intent analysis information included in the speech processing result. If there is no user account currently logged in to the server **400**, the electronic device **100** can determine that switching of the user account is necessary.

[0166] According to an embodiment, the electronic device **100** can determine whether switching of the user account is necessary based on the type of command corresponding to the speech. For example, the electronic device **100** can determine that switching of the user account is necessary based on the input speech corresponding to a command related to the user account (e.g., login, user account switching, or the like). For example, the electronic device **100** can determine that user account switching is necessary based on the input speech corresponding to a command for executing an application (e.g., OTT service, SNS, or the like) linked to the user account. For example, the electronic device **100** can determine that user account switching is necessary based on the input speech corresponding to a command (e.g., content search, content recommendation, speech recommendation, external device connection, or the like) using information on the user such as a usage history, a viewing history, and preferred genres.

[0167] In operation S**1504**, the electronic device **100** can switch the user account logged in to the server **400** when it is necessary to switch the currently logged-in user account. For example, the electronic device **100** can log out the currently logged-in first user account. Upon completion of logging out of the first user account, the electronic device **100** can log in to the server **400** with a second user account corresponding to the user identification information corresponding to the speech. The electronic device **100** can switch the user account after completion of generation and output of the first response.

[0168] The electronic device **100** can determine whether generation of the first response including a user indicator is necessary when the user who has uttered the speech corresponds to the currently logged-in user account or when switching of the user account is unnecessary in operation S**1505**. For example, the electronic device **100** can determine that generation of the first response is necessary when the input speech requires information on the user who has uttered the speech based on the intent analysis information included in the speech processing result.

[0169] The electronic device **100** can generate the first response including a user indicator in operation S**1506**. For example, the electronic device **100** can generate text (hereinafter referred to as response text) in response to the input speech based on the intent analysis information included in the speech processing result. Here, the electronic device **100** can generate text including both the response text and the user name as the first response.

[0170] The electronic device **100** can generate a second response that does not include a user indicator in operation S**1507**. The electronic device **100** can obtain information on the user in operation S**1508**. For example, the electronic device **100** can obtain a usage history stored in the storage **140** in relation to the user of the currently logged-in user account. For example, the electronic device **100** can obtain the name, nickname, icon, photo, age, sex, preferred genre, and the like stored in the server **400** in relation to the currently logged-in user account. According to an embodiment, the electronic device **100** can generate the first response or the second response after obtaining information on the user.

[0171] Referring back to FIG. **14**, the electronic device **100** can perform an operation according to the speech in operation S**1407**. For example, when the input speech requires user identification, the electronic device **100** can output either the first response or the second response. For example, the electronic device **100** can output the second response when the input speech does not require user identification. The electronic device **100** can perform an operation according to a command corresponding to the input speech based on the intent analysis information included in the speech processing result.

[0172] Referring to FIG. **16**, the electronic device **100** can output a first home screen **1600**. The

first home screen **1600** can be displayed while logged in to the server **400** with a first user account. Alternatively, the first home screen **1600** can be displayed when no user account is logged in to the server **400**.

[0173] Upon receiving speech uttered by a user, the electronic device **100** can output text **1610** corresponding to the speech uttered by the user based on the result of processing the speech received from the server **400**. The electronic device **100** can determine that the speech uttered by the user corresponds to a command related to the user account based on the speech processing result received from the server **400**.

[0174] Referring to reference numeral **1701** in FIG. **17**A, if a voice ID has not been registered in the user account of the user who has uttered the speech, user identification information corresponding to the speech processing result received from the server **400** may not be present. In this instance, the electronic device **100** can output a second response **1710** that requests registration of a voice ID and does not include a user indicator based on the fact that there is no user identification information corresponding to the speech uttered by the user.

[0175] Referring to reference numeral **1702** in FIG. **17**A, when a voice ID has been registered in the user account of the user who has uttered the speech and the use of the voice ID-related function is deactivated, the electronic device **100** can output a second response **1720** that does not include a user indicator and requests activation of use of the voice ID-related function.

[0176] Referring to reference numeral **1703** in FIG. **17**B, a voice ID can be registered in the user account of the user who has uttered the speech, and use of the voice ID-related function can be activated. Additionally, the user who has uttered the speech can correspond to the currently logged-in first user account. In this instance, the electronic device **100** can output a second response **1730** that does not include a user indicator and indicates that the user account of the user who has uttered the speech is currently logged in.

[0177] Referring to reference numeral **1704** in FIG. **17**B, a voice ID can be registered in the user account of the user who has uttered the speech, and use of the voice ID-related function can be activated. Additionally, the user who has uttered the speech may not correspond to the currently logged-in first user account. In this instance, the electronic device **100** can output a first response **1740** that includes a user indicator and indicates that login is performed with the user account of the user who has uttered the speech.

[0178] Referring to FIG. **18**, the electronic device **100** can switch the user account logged in to the server **400** from the currently logged-in first user account to a second user account of the user who has uttered the speech. For example, the electronic device **100** can log out of the currently logged-in first user account and then log in to the server **400** using the second user account of the user who has uttered the speech.

[0179] During switching between user accounts, the electronic device **100** can output a user switching screen **1800** related to user switching. The user switching screen **1800** can include an object **1810** indicating the first user account, an object **1820** indicating the second user account, and an object **1830** indicating user account switching.

[0180] Referring to FIG. **19**, the electronic device **100** can output a second home screen **1900** corresponding to the second user account while logged in with the second user account. The second home screen **1900** can be different from the first home screen **1600**. That is, the style of the home screen output from the electronic device **100**, for example, the types, number, and arrangement of objects included in the home screen, can be different for each user account.

[0181] Referring to FIG. **20**, the electronic device **100** can output the first home screen **1600**. Upon receiving speech uttered by a user, the electronic device **100** can output text **2010** corresponding to the speech uttered by the user based on the result of processing the speech received from the server **400**. The electronic device **100** can determine that the speech uttered by the user corresponds to a command for executing an exercise application based on the speech processing result received from the server **400**. Here, the exercise application can be an application linked to the user account.

[0182] Referring to reference numeral **2101** in FIG. **21**, the electronic device **100** can output a second response **2110** that does not include a user indicator and indicates execution of the exercise application. For example, if a voice ID has not been registered in the user account of the user who has uttered the speech, the user who has uttered the speech is not a user preset to use a voice ID, or the user who has uttered the speech corresponds to the currently logged-in first user account, the electronic device **100** can output a second response **2110** that does not include a user indicator. In this instance, the electronic device **100** can execute the exercise application without logging in to the user account for the server **400** or without user account switching.
[0183] Referring to reference numeral **2102** in FIG. **21**, the electronic device **100** can output a first response **2120** that includes a user indicator and indicates execution of the exercise application. For example, when the second user account is different from the currently logged-in first user account in a case in which use of the voice ID-related function is activated for the second user account of the user who has uttered the speech, the electronic device **100** can output the first response **2120** including a user indicator.
[0184] In this instance, the electronic device **100** can switch the user account logged in to the server **400** from the currently logged-in first user account to the second user account of the user who has uttered the speech. Additionally, the electronic device **100** can execute the exercise application while logged in with the second user account.
[0185] Referring to reference numeral **2201** in FIG. **22**, when an OTT service application is executed, the electronic device **100** can output an OTT service screen **2200** corresponding to an OTT service. Here, the OTT service application can be an application linked to the user account.
[0186] When the electronic device **100** receives speech uttered by a user while the OTT service screen **2200** is displayed, the electronic device **100** can output text **2210** corresponding to the speech uttered by the user based on the result of processing the speech received from the server **400**. The electronic device **100** can determine that the speech uttered by the user corresponds to a command for recommending video content based on the speech processing result received from the server **400**. Here, the command for recommending video content can be a command that uses information on the user.
[0187] Referring to reference numeral **2202** in FIG. **22**, the electronic device **100** can output a second response **2220** that does not include a user indicator and indicates recommendation of video content. For example, when a voice ID has not been registered in the user account of the user who has uttered the speech, the user who has uttered the speech is not a user preset to use a voice ID, or the user who has uttered the speech corresponds to the currently logged-in first user account, the electronic device **100** can output the second response **2220** that does not include a user indicator.
[0188] For example, even if use of the voice ID-related function is activated for the second user account of the user who has uttered the speech, the second user account is different from the currently logged-in first user account, and the uttered speech corresponds to the command for recommending video content, the electronic device **100** can determine that user account switching is unnecessary when an application linked to the user account is executed. That is, when an application linked to the user account is executed, it can be determined that user account switching is unnecessary, except for certain commands such as commands related to the user account or commands for executing applications linked to the user account.
[0189] Referring to FIG. **23**, the electronic device **100** can output the first home screen **1600**. Upon receiving speech uttered by a user, the electronic device **100** can output text **2310** corresponding to the speech uttered by the user based on the result of processing the speech received from the server **400**.
[0190] The electronic device **100** can determine that the speech uttered by the user corresponds to a command for recommending video content based on the speech processing result received from the server **400**. The command for recommending video content can be a command using information on the user.

[0191] Referring to reference numeral **2401** in FIG. **24**, the electronic device **100** can output a second response **2410** that does not include a user indicator and notifying of recommendation of video content. For example, if the user who has uttered the speech is a user preset to use a voice ID, and the user who has uttered the speech corresponds to the currently logged-in first user account, the electronic device **100** can output the second response including no user indicator.
[0192] The electronic device **100** can obtain information on the user of the first user account from the storage **140** and/or the server **400**. The electronic device **100** can output a recommended content list **2415** corresponding to the user of the first user account based on the information on the user obtained in response to the first user account. For example, if a history of viewing news content is stored in the storage **140** of the electronic device **100** or the preferred genre is set to news in the server **400** in relation to the user of the first user account, the recommended content list **2415** corresponding to the user of the first user account can include news content.
[0193] Referring to reference numeral **2402** in FIG. **24**, if use of the voice ID-related function is activated for the second user account of the user who has uttered the speech and the second user account is different from the currently logged-in first user account, the electronic device **100** can switch the user account logged in to the server **400** from the currently logged-in first user account to the second user account of the user who has uttered the speech. Additionally, the electronic device **100** can output a first response **2420** that includes a user indicator and notifies of recommendation of video content.
[0194] The electronic device **100** can obtain information on the user of the second user account from the storage **140** and/or the server **400**. The electronic device **100** can output a recommended content list **2425** corresponding to the user of the second user account based on the information on the user obtained in response to the second user account. For example, if a history of viewing sports content is stored in the storage **140** of the electronic device **100** or the preferred genre is set to sports in the server **400** in relation to the user of the second user account, the recommended content list **2425** corresponding to the user of the second user account can include sports content.
[0195] Referring to reference numeral **2501** in FIG. **25**, the electronic device **100** can output an OTT service screen **2200** corresponding to the OTT service. Upon receiving speech uttered by a user while the OTT service screen **2200** is displayed, the electronic device **100** can output text **2510** corresponding to the speech uttered by the user based on the result of processing the speech received from the server **400**. The electronic device **100** can determine that the speech uttered by the user is speech requesting information on the user who has uttered the speech based on the speech processing result received from the server **400**.
[0196] Referring to reference numeral **2502** in FIG. **25**, the electronic device **100** can output a first response **2520** that includes a user indicator and indicates the user who has uttered the speech. For example, the electronic device **100** can output the first response **2520** including a user indicator based on that fact that the user who has uttered the speech does not correspond to the currently logged-in user account, the command for identifying the user corresponds to a type of command that does not require user account switching, and the speech uttered by the user requests information on the user who has uttered the speech.
[0197] For example, when the user who has uttered the speech corresponds to the currently logged-in user account, the electronic device **100** can output the first response **2520** including the user indicator based on the fact that the speech uttered by the user requires information on the user who has uttered the speech.
[0198] Referring to FIG. **26**, the electronic device **100** can output the first home screen **1600**. Upon receiving speech uttered by a user, the electronic device **100** can output text **2610** corresponding to the speech uttered by the user based on the result of processing the speech received from the server **400**. The electronic device **100** can determine that the speech uttered by the user corresponds to a command for connecting a speaker among external devices based on the speech processing result received from the server **400**. The command for connecting an external device can be a command

using the information on the user.
[0199] Referring to reference numeral **2701** in FIG. **27**, the electronic device **100** can output a pop-up screen **2710** for a speaker that has been registered in the electronic device **100**. For example, when a voice ID has not been registered in the user account of the user who has uttered the speech, or use of the voice ID-related function is disabled for the user account of the user who has uttered the speech, the electronic device **100** can display the pop-up screen **2710** for a speaker that has been registered in the electronic device **100**.
[0200] For example, when use of the voice ID-related function is activated for the user account of the user who has uttered the speech, and at least one of speakers that have been registered in the electronic device **100** is set for public use, the electronic device **100 100** can display the pop-up screen **2710** for a speaker that has been registered in the electronic device **100**. Here, if the second user account of the user who has uttered the speech is different from the currently logged-in first user account, the electronic device **100** can switch the user account logged in to the server **400** from the currently logged-in first user account to the second user account of the user who has uttered the speech.
[0201] Referring to reference numeral **2702** in FIG. **27**, the electronic device **100** can output a pop-up screen **2720** for the speaker corresponding to the user who has uttered the speech. For example, when the use of the voice ID-related function is activated for the user account of the user who has uttered the speech, and all speakers that have been registered in the electronic device **100** are set for personal use, the electronic device **100** can display the pop-up screen **2720** for the speaker corresponding to the user who has uttered the speech. Here, if the second user account of the user who has uttered the speech is different from the currently logged-in first user account, the electronic device **100** can switch the user account logged in to the server **400** from the currently logged-in first user account to the second user account of the user who has uttered the speech.
[0202] Referring to FIG. **28**, the electronic device **100** can output the first home screen **1600**. Upon receiving speech uttered by a user, the electronic device **100** can output text **2810** corresponding to the speech uttered by the user based on the result of processing the speech received from the server **400**.
[0203] The electronic device **100** can determine that the speech uttered by the user corresponds to a command for connecting a terminal of the user who has uttered the speech among external devices based on the speech processing result received from the server **400**. The command for connecting an external device can be a command using the information on the user.
[0204] Referring to reference numeral **2901** in FIG. **29**, the electronic device **100** can output pop-up screens **2910** and **2920** for terminals that have been registered in the electronic device **100**. For example, when a voice ID has not been registered in the user account of the user who has uttered the speech, or the use of the voice ID-related function is disabled for the user account of the user who has uttered the speech, the electronic device **100** can display the pop-up screens **2910** and **2920** for terminals that have been registered in the electronic device **100**.
[0205] Referring to reference numeral **2902** in FIG. **29**, the electronic device **100** can output a screen **2900** corresponding to a terminal of the user who has uttered the speech. The screen **2900** corresponding to the terminal of the user who has uttered the speech can include an object **2930** corresponding to the screen of the terminal of the user who has uttered the speech.
[0206] For example, when the use of the voice ID-related function is activated for the user account of the user who has uttered the speech, and the user who has uttered the speech corresponds to the currently logged-in first user account, the electronic device **100** can display the screen **2900** corresponding to the terminal of the user who has uttered the speech without switching between user accounts for the server **400**.
[0207] For example, when the use of the voice ID-related function is activated for the second user account of the user who has uttered the speech, and the second user account is different from the currently logged-in first user account, the electronic device **100** can switch the user account logged

in to the server **400** from the currently logged-in first user account to the second user account of the user who has uttered the speech. Additionally, after logging in with the second user account, the electronic device **100** can display the screen **2900** corresponding to the terminal of the user who has uttered the speech.

[0208] As described above, according to at least one embodiment of the present disclosure, identification information on user voice can be registered in a user account. Additionally, according to at least one embodiment of the present disclosure, a user can be identified based on user voice.

[0209] It is possible to log in to an account of a user identified based on user voice. Additionally, according to at least one embodiment of the present disclosure, it is possible to accurately notify the user of switching of a logged-in user account.

[0210] Referring to FIGS. **1** to **29**, the electronic device **100** according to an aspect of the present disclosure includes the display **180**, the network interface **135** that communicates with the server **400**, the user input interface **150** that transmits a speech signal related to speech uttered by a user, and the controller **170**, wherein the controller **170** determines whether the user who has uttered the speech related to a user account logged in to the server **400**, outputs a first response including an indicator indicating the user who has uttered the speech through the display **180** based on the user who has uttered the speech being unrelated to the user account logged in to the server **400**, and outputs a second response not including the indicator through the display **180** based on the user who has uttered the speech being related to the user account logged in to the server **400**.

[0211] In addition, according to one aspect of the present disclosure, the electronic device further includes the memory **140** that stores a user list including at least one piece of user identification information related to a user account with a history of logging in to the server **400**, and the controller **170** can transmit the user list along with data including the speech signal to the server **400**, and determine whether the user who has uttered the speech is related to the user account logged in to the server **400** based on a result of processing the speech signal received from the server **400**.

[0212] In addition, according to one aspect of the present disclosure, the result of processing the speech signal can include first user identification information related to the speech uttered by the user, and the controller **170** can determine whether the user who has uttered the speech is related to the user account logged in to the server **400** by comparing the first user identification information with second user identification information related to the user account logged in to the server **400**.

[0213] In addition, according to one aspect of the present disclosure, the controller **170** can compare the first user identification information with the second user identification information based on being preset to use identification information related to speech with respect to the first user identification information, and generate the second response based on being preset not to use the identification information related to speech with respect to the first user identification information.

[0214] In addition, according to one aspect of the present disclosure, the controller **170** can log out of the user account logged in to the server **400** based on the user who has uttered the speech being unrelated to the user account logged in to the server **400** and log in to the user account related to the user who has uttered the speech.

[0215] In addition, according to one aspect of the present disclosure, the controller **170** can determine whether the speech uttered by the user requires identification of the user based on a result of performing natural language processing on the speech signal, determine whether the user who has uttered the speech is related to the user account logged in to the server **400** based on the speech uttered by the user requiring identification of the user, and generate the second response based on the speech uttered by the user not requiring identification of the user.

[0216] In addition, according to one aspect of the present disclosure, the electronic device **100** further includes the memory **140** that stores user histories of using the electronic device **100**, and the controller **170** can determine that the speech uttered by the user requires identification of the

user based on the speech uttered by the user being related to a command using a history stored in the memory, and determine that the speech uttered by the user does not require identification of the user based on the speech uttered by the user being unrelated to the command using a history stored in the memory **140**.

[0217] In addition, according to one aspect of the present disclosure, the controller **170** can determine that the speech uttered by the user requires identification of the user based on the speech uttered by the user being related to a first application linked to the user account, and determine that the speech uttered by the user does not require identification of the user based on the speech uttered by the user being related to a second application not linked to the user account.

[0218] In addition, according to one aspect of the present disclosure, the controller **170** can determine whether the user who has uttered the speech is related to the user account logged in to the server **400** based on the speech uttered by the user being related to a first command requiring identification of the user and related to the user account while an application linked to the user account is executed, and generate the second response based on the speech uttered by the user being related to a second command different from the first command and requiring identification of the user.

[0219] Additionally, according to one aspect of the present disclosure, the indicator can include at least one of a name, a nickname, an icon, or a photo previously stored with respect to the user who has uttered the speech.

[0220] The system **10** according to one aspect of the present disclosure includes the electronic device **100** and the server **400**, and the electronic device **100** transmits data including a speech signal related to speech uttered by a user to the server **400**, determines whether the user who has uttered the speech is related to a user account logged in to the server **400** based on a result of processing the speech signal received from the server **400**, outputs a first response including an indicator indicating the user who has uttered the speech through the display **160** based on the user who has uttered the speech being unrelated to the user account logged in to the server **400**, and outputs a second response not including the indicator through the display **180** based on the user who has uttered the speech being related to the user account logged into the server **400**, and the server **400** can generate identification information for the speech signal included in the data received from the electronic device **100**, determine first identification information related to the identification information for the speech signal among identification information mapped to user identification information related to user accounts stored in the database **490** of the server **400**, and transmit the result of processing the speech signal including first user identification information mapped to the first identification information to the electronic device **100**.

[0221] Additionally, according to one aspect of the present disclosure, the electronic device **100** can store a user list including at least one user identification information related to a user account with a history of logging in to the server **400**, and transmit the user list to the server **400** along with the data including the speech signal, and the server **400** can search the database **490** for user identification information related to the user list, and determine whether identification information mapped to the user identification information related to the user list is related to the identification information for the speech signal.

[0222] Additionally, according to one aspect of the present disclosure, the electronic device **100** can determine whether the user who has uttered the speech is related to the user account logged in to the server **400** by comparing the first user identification information with second user identification information related to the user account logged in to the server **400**.

[0223] Additionally, according to one aspect of the present disclosure, the electronic device **100** can log out of the user account logged in to the server **400** based on the user who has uttered the speech being unrelated to the user account logged in to the server **400** and log in to the user account related to the user who has uttered the speech.

[0224] In addition, according to one aspect of the present disclosure, the server **400** can transmit a

result of processing the speech signal, including a result of performing natural language processing on the speech signal, to the electronic device **100**, and the electronic device **100** can determine whether the speech uttered by the user requires user identification based on the result of performing natural language processing, determine whether the user who has uttered the speech is related to the user account logged in to the server **400** based on the speech uttered by the user requiring identification of the user, and generate the second response based on the speech uttered by the user not requiring identification of the user.

[0225] 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 spirit and technical scope of the present disclosure are not limited thereby.

[0226] Meanwhile, the operating method of the present disclosure can 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 can 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 can be stored and executed in a distributed manner.

[0227] The effects of the electronic device and the system including the same according to the present disclosure will be described below.

[0228] According to at least one embodiment of the present disclosure, it is possible to register identification information on user voice in a user account. According to at least one embodiment of the present disclosure, it is possible to identify a user based on user voice. According to at least one embodiment of the present disclosure, it is possible to log in to a user account identified based on user voice. According to at least one embodiment of the present disclosure, it is possible to accurately notify a user of switching of a logged-in user account.

[0229] 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 spirit and scope of the present disclosure can 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.

[0230] 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 and spirit of the disclosure as disclosed in the accompanying claims.

Claims

1. An electronic device comprising: a display; a network interface configured to communicate with a server; a user input interface configured to receive speech uttered by a user; and a controller configured to: determine whether the speech uttered by the user is related to a user account logged in to the server, in response to determining the speech uttered by the user is unrelated to the user account logged in to the server, display a first response on the display including an indicator indicating the user who has uttered the speech is not the user logged in to the user account, and in response to determining the speech uttered by the user is related to the user account logged in to the server, display a second response on the display indicating the user who uttered the speech is currently logged in to the server.

2. The electronic device of claim 1, further comprising: a memory configured to store a user list including at least one piece of user identification information related to a user account with a

history of logging in to the server, wherein the controller is further configured to: transmit the user list to the server along with data including the speech signal, and determine the speech uttered by the user is related to the user account logged in to the server based on a result of processing the speech signal received from the server.

3. The electronic device of claim 2, wherein the result of processing the speech signal includes first user identification information related to the speech uttered by the user, and wherein the controller is further configured to determine the speech uttered by the user is related to the user account logged in to the server by comparing the first user identification information with second user identification information related to the user account logged in to the server.

4. The electronic device of claim 3, wherein the controller is further configured to: compare the first user identification information with the second user identification information based on the electronic device being preset to use voice-related identification information with respect to the first user identification information, and generate the second response based on the electronic device being not preset to use the voice-related identification information with respect to the first user identification information.

5. The electronic device of claim 1, wherein the controller is further configured to: log out of the user account logged in to the server based on the speech uttered by the user being unrelated to the user account logged in to the server, and log in to a user account related to the user who has uttered the speech.

6. The electronic device of claim 1, wherein the controller is further configured to: determine whether the speech uttered by the user requires identification of the user based on a result of performing natural language processing on the speech uttered by the user, determine whether the speech uttered by the user is related to the user account logged in to the server based on the speech uttered by the user requiring the identification of the user, and generate the second response based on the speech uttered by the user not requiring the identification of the user.

7. The electronic device of claim 6, further comprising: a memory configured to store user histories of users using the electronic device, wherein the controller is further configured to: determine the speech uttered by the user requires user identification based on the speech uttered by the user being related to a command the history stored in the memory, and determine the speech uttered by the user does not require the user identification based on the speech uttered by the user being unrelated to the command using the history stored in the memory.

8. The electronic device of claim 6, wherein the controller is further configured to: determine the speech uttered by the user requires user identification based on the speech uttered by the user being related to a first application linked to the user account, and determine the speech uttered by the user does not require the user identification of the user based on the speech uttered by the user being related to a second application not linked to the user account.

9. The electronic device of claim 6, wherein the controller is further configured to: execute an application linked to the user account, determine whether the speech uttered by the user is related to the user account logged in to the server based on the speech uttered by the user being related to a first command requiring the user identification and related to the user account, and generate the second response based on the speech uttered by the user being related to a second command different from the first command and requiring the user identification.

10. The electronic device of claim 1, wherein the indicator includes at least one of a name, a nickname, an icon, or a photo previously stored with respect to the user who has uttered the speech.

11. A system comprising: a server; and an electronic device configured to: receive speech uttered by a user, transmit data including a speech signal related to the speed uttered by the user to the server, receive information from the server indicating whether the speech uttered by the user is related to a user account logged in to the server, in response to the speech uttered by the user being unrelated to the user account logged in to the server, display a first response on the display including an indicator indicating the user who has uttered the speech is not the user logged in to the

user account, and in response to the speech uttered by the user is related to the user account logged in to the server, display a second response on the display indicating the user who uttered the speech is currently logged in to the server, wherein the server is configured to: generate identification information for the speech signal included in the data transmitted by the electronic device, determine first identification information related to the identification information for the speech signal among identification information mapped to user identification information related to user accounts stored in a database of the server, and transmit the result of processing the speech signal including first user identification information mapped to the first identification information to the electronic device.

12. The system of claim 11, wherein the electronic device is configured to: store a user list including at least one piece of user identification information related to a user account with a history of logging in to the server, and transmit the user list to the server along with the data including the speech signal, and wherein the server is configured to: search the database for user identification information related to the user list, and determine whether identification information mapped to the user identification information related to the user list is related to the identification information for the speech signal.

13. The system of claim 11, wherein the electronic device is further configured to: determine whether the speech uttered by the user is related to the user account logged in to the server by comparing the first user identification information with second user identification information related to the user account logged in to the server.

14. The system of claim 11, wherein the electronic device is further configured to: log out of the user account logged in to the server based on the speech uttered by the user being unrelated to the user account logged in to the server, and log in to the user account related to the user who has uttered the speech.

15. The system of claim 11, wherein the server is further configured to: transmit a result of processing the speech signal, including a result of performing natural language processing on the speech signal, to the electronic device, and wherein the electronic device is further configured to: determine whether the speech uttered by the user requires user identification based on the result of performing natural language processing, determine whether the speech uttered by the user is related to the user account logged in to the server based on the speech uttered by the user requiring identification of the user, and generate the second response based on the speech uttered by the user not requiring identification of the user.