- An Interactive Answer Display Terminal Case -

1. Case

We assume that the inventive step of a patented invention is disputed in a Trial for Patent Invalidation.

2. Background information

Previously, users have received services from business operators by visiting stores or engaging in direct oral conversations with business operators through phone calls, etc. Nowadays, with the development of communication networks, users can receive services via communication networks. Internet banking is a typical example, and users can conduct various commercial transactions anytime, anywhere by operating communication terminals.

Recently, voice recognition technology has been developed that is equipped with artificial intelligence, and support systems for operator work which has been dealt with in business operators, such as responding to requests from users and handling complaints, have also been created. For example, support systems handling users more appropriately and promptly have been developed by extracting the intention of user's requests, questions, or complaints based on analyses of voice information input within the communication terminal by users.

3. Patented Invention

Date of filing   February 25, 2016

Claim

"Information providing system including a server and a communication terminal: wherein, said server transmits an answer message for a user's voice information transmitted from a communication terminal, or an inquiry message if the answer message cannot be specified, to the communication terminal;

said communication terminal displays a virtual operator for imitating an employee who handles users;

wherein the virtual operator is displayed so that a part of the virtual operator has larger movement when the answer message or the inquiry message is played than when the answer message or the inquiry message is not played."
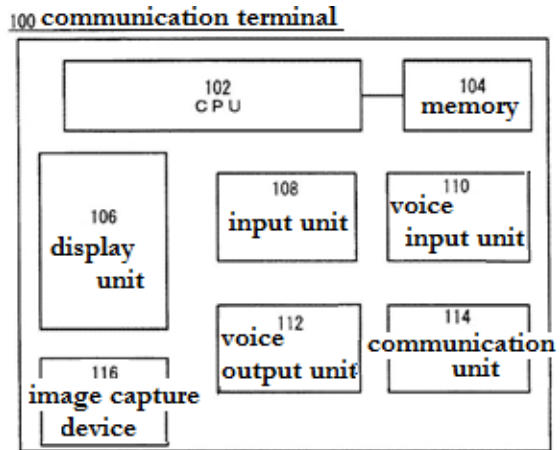
Description and Drawings

[Problem to be solved]

The objective of the present invention is to reduce a feeling of distance between users and business operators, and to build a closer user-business operator relationship through natural communication means via a mobile terminal.

[Embodiment]

[Fig. 1]

Fig. 1 illustrates a system diagram of a communication terminal 100 that can be used by the information-providing method of this embodiment. The communication terminal 100 is a terminal that can be operated by a user, and examples thereof include mobile communication terminals such as mobile phones, tablets, and smartphones.

[Fig. 1]
100 communication terminal
102 CPU
104 memory
106 display unit
108 input unit
110 voice input unit
112 voice output unit
114 communication unit
116 image capture device

[Fig. 2]

[Fig. 2]
I answer questions regarding our various procedures. Please tell me your question.
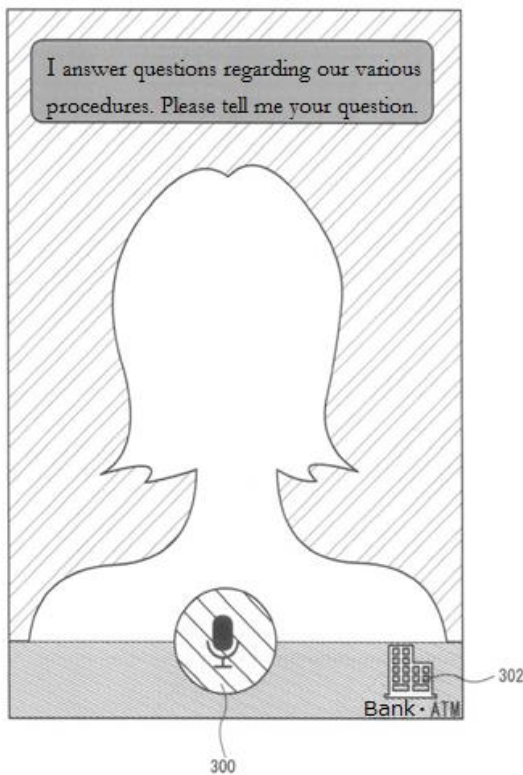Bank·ATM 302
300

Fig. 2 is an example of a virtual operator, and illustrates a display state on a display unit 106 of the communication terminal 100. The virtual operator may imitate an employee who handles users. In addition, the virtual operator may be programmed to perform natural human movements (for example, changes in facial expressions, movements of the head and hands, etc.) to give the user a sense of closeness.

Fig. 3 shows a flowchart of the information-providing method of this embodiment. When a user starts the application (step 200), the program displays a virtual operator on the display unit 106 of the communication terminal (step 202). The user first operates the icon 300 to activate a voice input unit 110. After this, the user verbally asks a question to the business operator (step 210).

Based on the instruction of the program, the voice input unit 110 transmits the user's question as voice information to a communication unit 114, and the communication unit 114 transmits the voice information to the API which

[Fig. 3]

is the interface of the voice recognition server via the communication network (step 212). The transmitted voice information is analyzed by a voice recognition server and converted into text data (step 214). The voice information converted into text data is analyzed by a dialogue server, and the intention of the user's question is understood and grasped (step 216).

After understanding the intention of the user's question, the dialogue server narrows down answer messages using a narrowing engine (step 224). Then, a suitable answer message is selected from various answer messages stored in the database of the dialogue server (step 226). The text information of the selected answer message is converted into voice information by a voice synthesis server, and transmitted to the user's communication terminal 100 through the API of the voice synthesis server (step 226). At this time, the answer message may be transmitted to the communication terminal 100 as not only text information, but also voice information. According to the instruction of the program, the voice output unit 112 plays the voice information received by the communication terminal
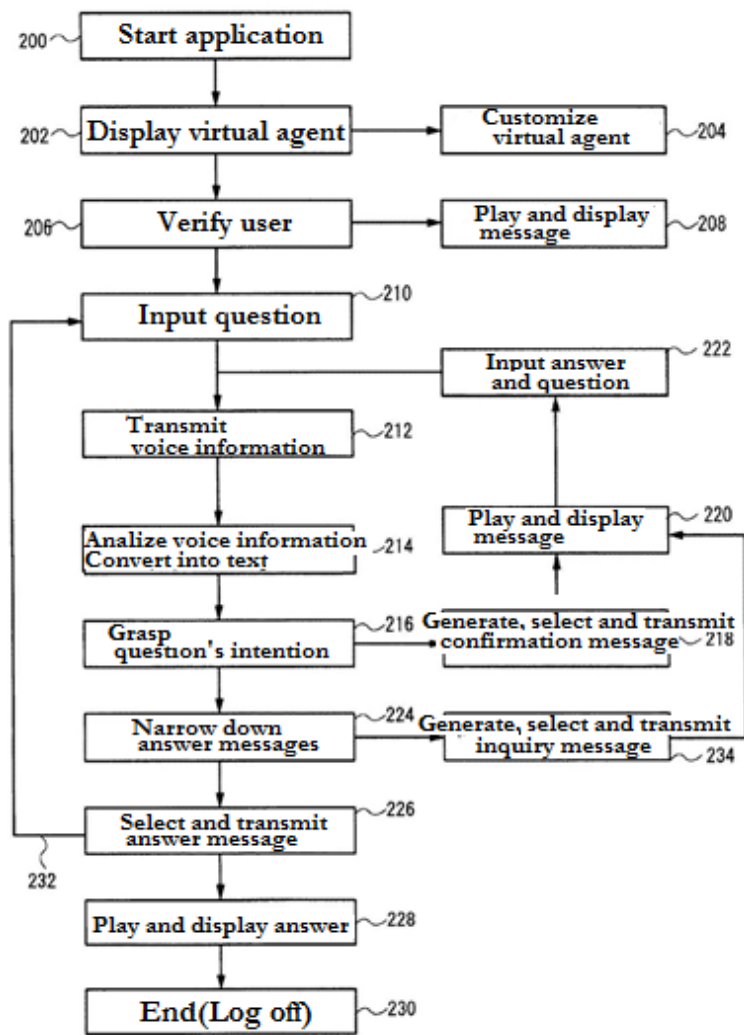
100 (step 228). At the same time, the answer message may be displayed as text on the display unit 106.

Standard messages created by business operators may be stored in the database as answer messages, and the dialogue server can select an answer message from the stored messages. However, the information-providing method of this embodiment is not limited to this. In addition, not only the answer message is selected, but also the function as artificial intelligence may be installed in the dialogue server, and new answer messages created through learning using dialogue with the user may be converted into voice information by the voice synthesis server and then transmitted to the communication terminal 100. That is, the dialogue server may be provided with a function of obtaining new conclusions using dialogue with users and a function of learning using dialogue history with users, and answer messages may be created by using these functions.

In step 224, when answer messages for a user's question cannot be narrowed down, or when a large number of answer messages must be selected, the user may be requested to provide further information through answering a generated or selected question (inquiry message) transmitted to the communication terminal 100 for the user (step 234). The inquiry message is played or/and displayed on the communication terminal 100 as voice, text, or both. The user can respond to the inquiry message by voice or text (step 222). By repeating these steps of transmitting and responding to an inquiry message, it becomes possible to find out the answer and information requested by the user from a large amount of information that business operators can provide, through the dialogue between the user and the virtual operator.

Further, when the above-mentioned various messages are played by the voice output unit 112, the mouth and eyes of the virtual operator may be moved in order to make the virtual operator appear to be actually speaking. Or, the virtual operator may make an explanatory gesture, such as moving the hands. That is, it may be programmed so that a part of the virtual operators has larger movement than when the message is not played.

[Effect]
As described above, in this embodiment, users can transmit various questions to business operators through dialogues between virtual operators and users. Therefore, users are free from entering texts using a keyboard or touch panel, and it becomes possible to more easily transmit various questions to business operators. In addition, since it uses voice synthesis technology to interact with the virtual operators, it can provide natural dialogue and build a closer user-business operator relationship.

4. Document 1 (D1) (Published Patent Application)

Date of publication   February 8, 2016

D1 Invention

"An interactive processing system including a server unit and a user terminal device: wherein the user terminal device transmits a user voice question to the server unit, the server unit transmits an expected answer or an inquiry question corresponding to the user voice question to the user terminal device, and the user terminal device outputs by voice and displays the expected answer or the inquiry question."

Description and Drawings

[Background Art]

Voice dialogue devices have already been provided wherein computers search for an answer content to a user's voice information, and return the answer content to the user. However, in the conventional technology, when the voice information is ambiguous, there is a risk that voice dialogue devices will return inappropriate answer content to the user.

[Problem to be solved]

The present invention provides an interactive processing system for narrowing down a user's question to a more specific question content by asking the user a question when a plurality of answer contents can be applied.

[Embodiment]

Fig. 1 is a diagram of an interactive processing system according to this embodiment. As shown in Fig. 1, the interactive processing system 1 according to the present embodiment exemplifies a user terminal device 10, a server management device 12, a voice recognition processing server 14, a language processing server 18, and a voice synthesis processing server 16. This interactive processing system 1 functions to realize an interactive FAQ system with the above configuration. In short, this

[Fig. 1]

voice recognition processing server — 14

user terminal device — 10

server management device — 12

language processing server — 18

voice synthesis processing server — 16

1

is the system in which the answer content is returned by the process of the voice
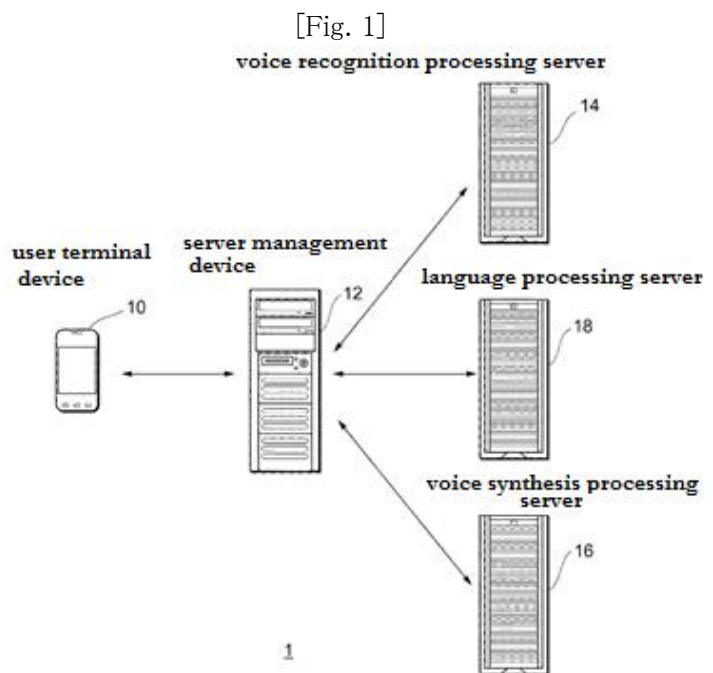
recognition processing server 14, the language processing server 18, and the voice synthesis processing server 16 when the user asks the user terminal device 10 by voice.

The user terminal device 10 is a mobile phone such as a smartphone, tablet terminal, or the like, and has a communication function for transmitting and receiving user information, voice information (user question), and character string information.

Fig. 2 is a block diagram showing an example of the functional configuration of the interactive processing system 1. As shown in Fig. 2, the interactive processing system 1 includes the user terminal device 10 and a server unit 20. The user terminal device 10 includes an output unit 26 and an operation display unit 28. The output unit 26 converts the answer content into voice and outputs it from a speaker, and the operation display unit 28 displays the answer content. Further, in the server unit 20, the search unit 46 searches all or a part of the plurality of contents (plurality of answer contents) in the content table 58 based on the user character string question received by the second reception unit 44, and acquires either the expected answer from multiple answer contents or the inquiry question, corresponding to the user character string question. Specifically, the search unit 46 searches the content table 58 to calculate the correlations between the user question character string and each content by a predetermined calculation formula, compares them with each other, and specifies a certain content. If the certain content can be specified, it means that the user character string question is specific enough to answer, and the answer content recorded in the content is acquired.

[Fig. 2]
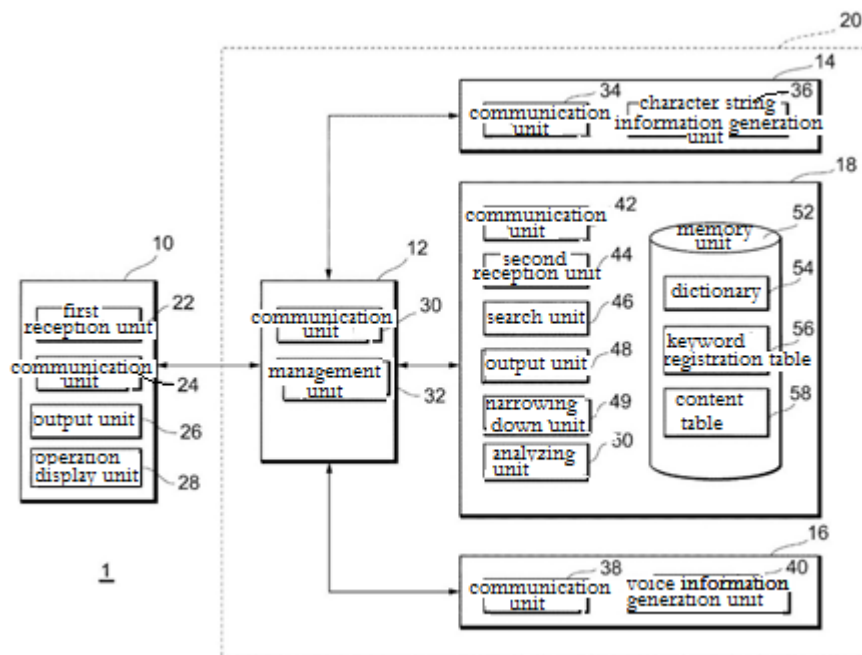
Alternatively, it implies that the user character string question is ambiguous when a certain content cannot be specified because a plurality of contents may be applicable even if the content table 58 is searched, so that the search unit 46 further searches to specify the content in which the inquiry question for narrowing down the ambiguous user question is stored, and to acquire the inquiry question recorded in the content.

For example, as shown in Fig. 3, it is assumed that the user asks a question by voice (user voice question) from the user terminal device 10 saying, "Tell me about mobile phone A." The voice recognition server 14 that has input the user voice question executes the voice recognition process, and transmits the user character string question obtained by converting the user voice question into the corresponding character string, to the language processing server 18. In the language processing server 18, the search unit 46 acquires an inquiry content as an inquiry question recorded in the specified content 59 in the content table 58 corresponding to the user character string question: for example, "What do you want to know about mobile phone A?" The inquiry content is then transmitted to the voice synthesis server 16, and the voice synthesis server 16 performs voice synthesis processing, and outputs the voice corresponding to the character string to the user terminal device 10.

[Fig. 3]
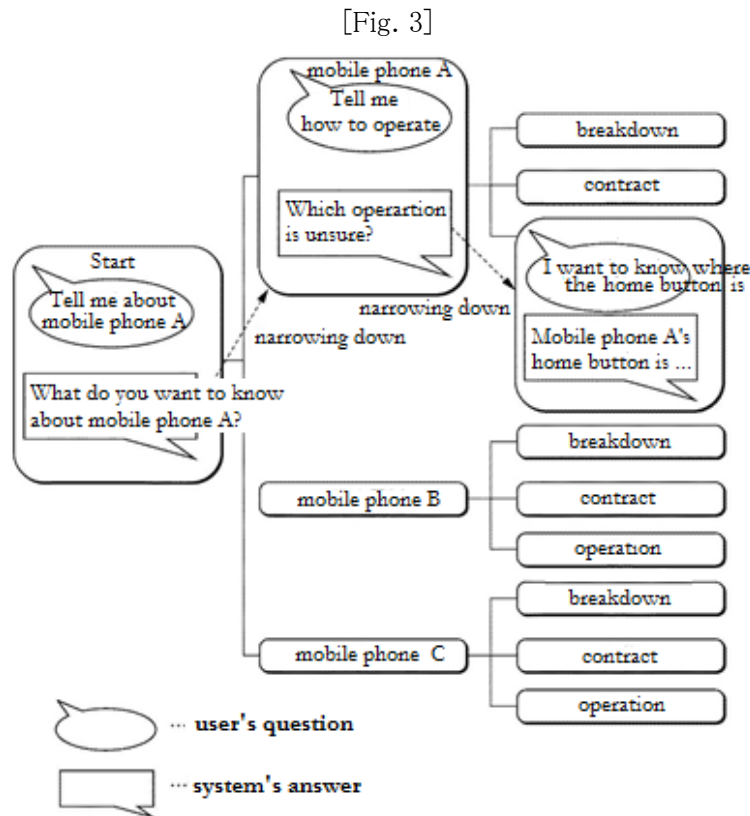
In response, when a further user voice question such as "Tell me how to operate." is input, and the corresponding user character string question is transmitted from the voice recognition server 14, the search unit 46 narrows down contents corresponding to the user

question. Then, the content in which the "operation method" included in the user question is recorded is specified as the record having the highest correlation with the user question, and the inquiry content recorded in the content "which operation method is unsure?" is acquired and output.

In response, when a further user voice question such as "I want to know where the home button is" is input and the corresponding user character string question is transmitted from the voice recognition server 14, the search unit 46 specifies the content in which the "operation" is classified corresponding to the user question, and the recorded answer content "Mobile phone A's home button is..." is acquired and output.

[Effect]
This invention provides an interactive processing system which can return an appropriate answer content to a user's question, including ambiguous content, by narrowing down answer contents with repeated user's questions and inquiry questions.

5. Document 2 (D2) (Published Patent Application)

Date of publication   February 12, 2015
D2 Technology
"A navigation device for displaying an agent on a display:
wherein the agent is displayed so that the agent's mouth is opened when an answer voice data of a responding center is output from a speaker as compared with the case of waiting in order to display as if the agent is speaking."

Description and Drawings
[Background Art]
In answer systems including a navigation device and an information center for wireless communication with the navigation device, the navigation device acquires data for a user's voice and transmits it to the information center. The information center then performs voice recognition processing on the voice data and returns an answer voice data based on the recognition processing result to the navigation device. Then, the navigation device receiving the answer voice data outputs the voice based on the answer voice data.

On the other hand, also in mobile phones, answer systems for providing various information based on a user's voice input have become widespread via a wireless communication between the information center managed by a mobile phone company and mobile phones.

In recent years, a technology for operating navigation devices and mobile phones in cooperation with each other have also been developed. In this way, when navigation devices and mobile phones are operated in cooperation with each other, a plurality of answer systems can be used via the navigation device.

[Problem to be solved]
In such a configuration, however, the user must select an answer system according to the purpose from a plurality of answer systems by operating a navigation device or the like, and then input the voice in the answer system. If the user selects an answer system but then hopes to use a different one, the user must perform an operation to switch answer systems.

The present invention is invented based on this understanding, and an objective of the present invention is to provide an answer system which can save a user the hassle of selecting an answer system when a plurality of answer systems responding to a user's voice input are available.

[Embodiment]
As shown in Fig. 1, an answer control system 100 includes a navigation device 1, a mobile phone 2, a first center 3, and a second center 4. Each pair of the navigation device 1 and the first center 3, the navigation device 1 and the mobile phone 2, and the mobile phone 2 and the second center 4 transmit and receive data using known wireless communication techniques.

[Fig. 1]
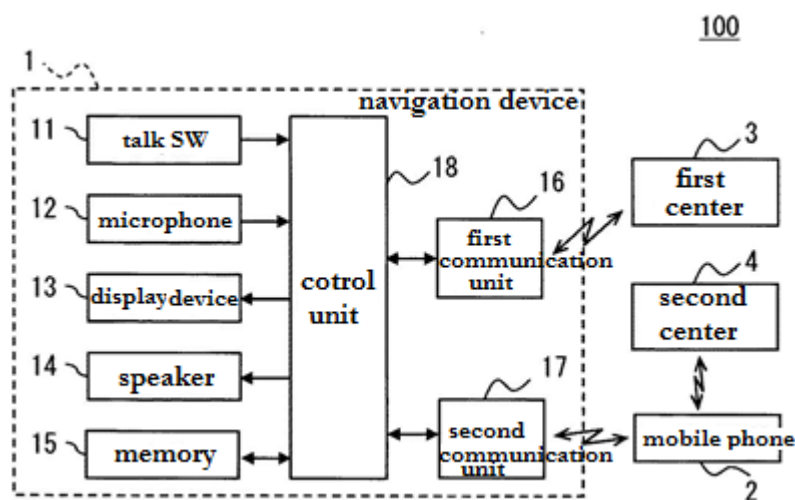
The first center 3 is, for example, an information center of an automobile company, and responds to user's questions related to the operation of a navigation device, traffic jam
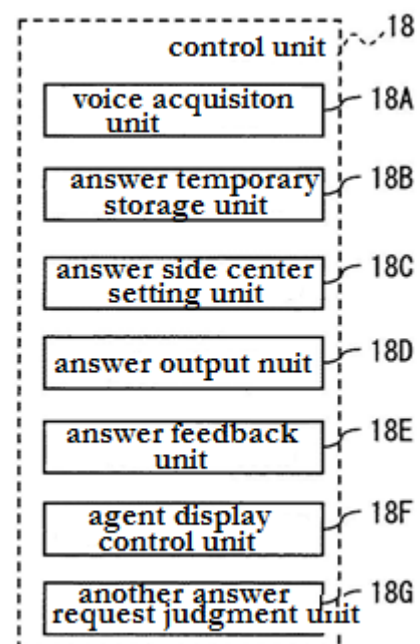
9

information, operation of the own vehicle, and also responds to commands by voice input to use the function of navigation device 1. As an outline of the operation of the first center 3, the user's question content is analyzed by performing voice recognition processing based on the voice data transmitted from the navigation device 1. Then, it is determined which center should respond to the user's question, and an answer voice data for the user's question is generated and returned to the navigation device 1. For example, the first center 3 tells the location of the nearest convenience store from the user's current location in response to the user's question "Where is the convenience store?".

The second center 4 is, for example, an information center of a mobile phone company, and provides various services to users of the mobile phone 2 which use the mobile phone network managed by the mobile phone company. The second center 4 has the same configuration as the first center 3, except that it does not determine which center should respond to the user's question. That is, the user's question content is analyzed by performing voice recognition processing based on the voice data transmitted from the navigation device 1. Then, the answer voice data that is the response to the user's question is generated and returned to the navigation device 1. For example, assuming a schedule management function as a service provided by a mobile phone company, the second center 4 tells you the schedule for the day to respond to the user's question "What is my schedule for today?"

[Fig. 2]

As shown in Fig. 2, a control unit 18 includes a functional block for executing various processes. An answer center setting unit 18C decides which of the first center 3 and the second center 4 should output the answer voice data from a speaker 14 as an answer to the user's question input by the user based on the judgment result signal sent from the first center 3.
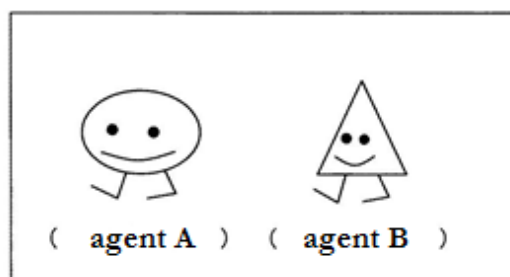
When each answer system becomes available, the agent display control unit 18F simultaneously displays agents A and B corresponding to each answer system, as shown in Fig. 3 (A). Agent A in Fig. 3 is an image of an agent corresponding to the answer system by the first center 3, and agent B is an image of an agent corresponding to the answer system by the second center 4. These agents are simplified characters such as animals.

control unit — 18
- voice acquisiton unit — 18A
- answer temporary storage unit — 18B
- answer side center setting unit — 18C
- answer output nuit — 18D
- answer feedback unit — 18E
- agent display control unit — 18F
- another answer request judgment unit — 18G
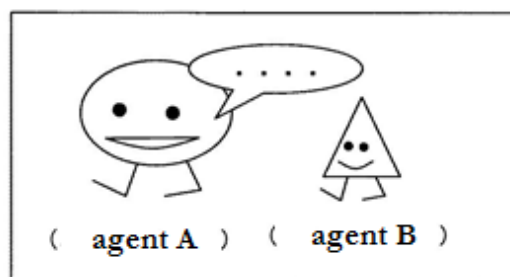
The agent display control unit 18F generates agent images corresponding to each answer system using a known technique, and displays the images of the plurality of agents by synthesizing or superimposing them so as to be displayed on one screen.
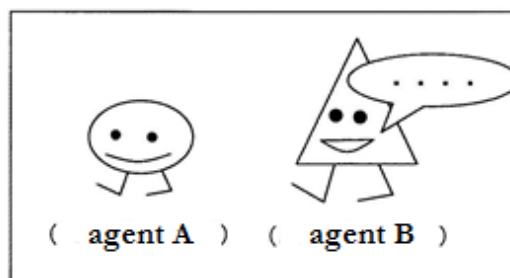
When the answer output unit 18D outputs the answer voice data from the answer side center to the speaker 14 by the answer side center setting unit 18C, the agent corresponding to the answer side center is displayed to be relatively large and open the mouth as if it is speaking, as shown in Fig. 3 (B) and (C).

[Effect]
According to the above configuration, the center (that is, the answer system) to respond to the user's voice input is automatically selected based on the content of the input voice data, so that the system can save the user the hassle of selecting—thereby being more convenient for the user.

[Fig. 3]

(A) display example in waiting (S103)

(B) display example when first center answers

(C) display example when second center answers

Further, the agent display control unit displays the agents A and B corresponding to each center on the display device 13, and the agent corresponding to the center that generated the answer voice data output from the speaker 14 by the answer output unit 18D is displayed so that the agent opens the mouth as if it is speaking. As a result, the user can recognize at a glance which center corresponds to the voice input.

6. Document 3 (D3) (Published Patent Application)

Date of publication   July 8, 2010

Well-known matter exemplified in D3[1]

> "In a voice dialogue system, a person who imitates an employee who handles users is displayed as a still image."

Description and Drawings

Fig. 1 is a system diagram showing a schematic overall configuration of a visitor reception system 1 of this embodiment. In Fig. 1, the visitor reception system 1 is, for example, a system that performs reception work for visitors to buildings, companies, and other architectures. In this example, the case where it is installed in a company will be described as an example.

[Fig. 1]
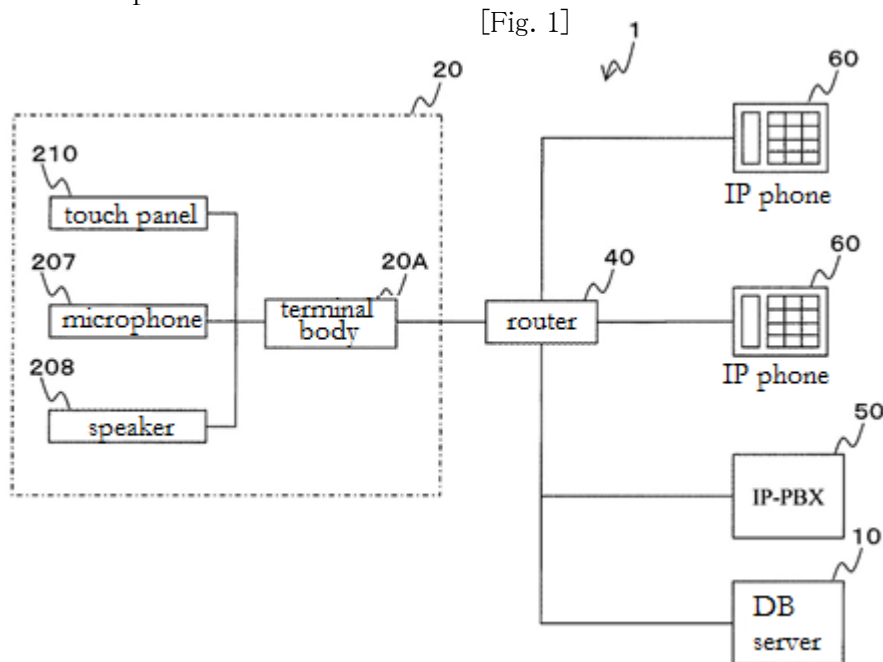
The visitor reception system 1 includes, for example, a reception terminal 20 (visitor reception device) installed near the entrance of the company, a DB server 10 composed of a well-known personal computer, a plurality of IP telephones 60 for each company's employee, and an IP-PBX 50 which is a well-known exchange device for exchanging lines of the plurality of IP telephones 60—all of which are connected via a router 40.

The reception terminal 20 has a terminal body 20A and a touch panel 210, a microphone 207 and a speaker 208 connected to the terminal body 20A.

---

[1] The well-known matter is described in many documents other than D3, and there is no contention between the parties regarding the well-known matter.

The speaker 208 functions as a notification means for notifying the visitor by outputting the voice into which is converted the voice signal input from the terminal body 20A. The microphone 207 functions as a voice input means, converts the voice input by the visitor into voice information, and outputs it to the terminal body 20A.

Fig. 2 is a perspective view showing an example of the detailed external structure of the touch panel 210 and the microphone 207 provided in the reception terminal 20.

[Fig. 2]

In Fig. 2, the touch panel 210 is supported via an arm 211 with respect to the base 212 installed horizontally in this example, and the surface direction is diagonally upward so as to be perpendicular to the line of sight of the operator (that is, the visitor in this example). The microphone 207 is arranged in a substantial arc shape with the tip of the microphone 207 facing the visitor side with respect to the base 212.
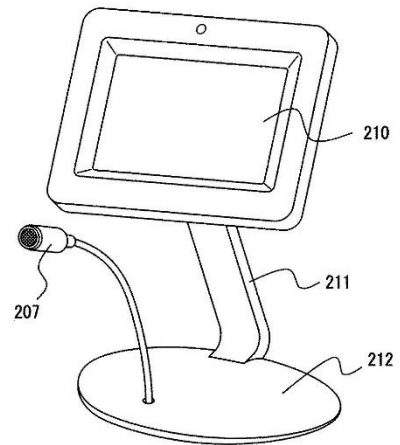
Fig. 3 is a drawing showing an example of a display screen on the touch panel 210. On this screen, the virtual person M who performs the reception work generated by the drawing program is displayed as a still image together with the office-like background G. In addition, a sentence L (abbreviated as "***" in the figure) corresponding to the voice uttered from the speaker 208 is also displayed.

[Fig. 3]

13
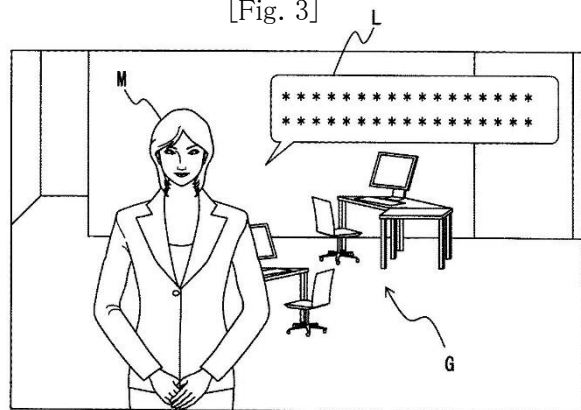
7. Key points of each Published Document

The key points of documents 1 to 3 are as follows:

Document 1 (D1)

- It is an invention of an interactive processing system that responds to a user's voice question only by text and voice.
- A terminal device is a mobile phone such as a smartphone, tablet terminal, or the like.
- D1 does not disclose the problem to be solved by the patented invention (to reduce a feeling of distance between users and business operators and build a closer user-business relationship).
- Displaying a virtual operator is not disclosed.

Document 2 (D2)

- It is premised that a plurality of answer systems responding to a user's voice input are available via a communication terminal.
- The communication terminal is a navigation device.
- D2 does not disclose the problem to be solved by the patented invention.
- It is disclosed that an agent image corresponding to each answer system is displayed on one screen while the plurality of answer systems are available. These agents are simplified characters such as animals, and the agent corresponding to the answer system is displayed when answering so that it opens the mouth as if it is speaking when the voice data is output from a speaker.

Document 3 (D3)

- This is an example showing that it is a well-known matter to display a person imitating an employee who handles users as a still image in a voice dialogue system.
- A communication terminal is a reception terminal (dedicated fixed terminal) of a visitor reception system.
- D3 does not disclose the problem to be solved by the patented invention.
- A person who imitates an employee who handles users is displayed as a still image.

8. Identical features and differences

The identical features and differences between the patented invention and D1 invention are as follows:

Identical features

"Information providing system including a server and a communication terminal: wherein, said server transmits an answer message for a user's voice information transmitted from a communication terminal, or an inquiry message if the answer message cannot be specified, to the communication terminal;
said communication terminal plays the answer message or the inquiry message."

Differences

The patented invention displays "a virtual operator for imitating an employee who handles users" on the communication terminal, whereas the D1 invention does not display a virtual operator. Furthermore, the patented invention states "the virtual operator is displayed so that a part of the virtual operator has larger movement when the answer message or the inquiry message is played than when the answer message or the inquiry message is not played", whereas the D1 invention does not disclose the specific matter.

There is no contention between the parties regarding the identical features and differences.

9. Discussion Points

（1）Problem to be solved

[Discussion Point] Is it reasonable to recognize an issue which is not explicitly disclosed in D1 and D2, such as "to facilitate media communication", as a well-known issue that does not need to exemplify literature[2], and motivate it to apply D2 technology to the D1 invention?   Reasonable：Y   Unreasonable：N

（2）Teaching away

[Discussion Point] Is it reasonable to apply D2 technology, which is premised on multiple answer systems, to the D1 invention of a single interactive processing system?[3]   Reasonable：Y   Unreasonable：N

（3）Application of secondary reference and well-known matter to primary reference

[Discussion Point] Is it reasonable to deny the inventive step of the patented invention by applying D2 technology and the well-known matter exemplified in D3 to the D1 invention?[4]   Reasonable：Y   Unreasonable：N

（4）Conclusion

[Discussion Point] Considering the above (1) to (3) in a comprehensive manner, is the patented invention invalid?   Invalid：Y   Not invalid（Valid）：N

[End of file]

---

[2] Between the parties, there is no contention that the issue is well known and doesn't need to exemplify literature.

[3] The patentee insists, "In D2, the reason why the agent is displayed to open the mouth is that users can recognize at a glance which center corresponds to the user's voice input. Moreover, from the problem to be solved in D2 (to provide an answer system which can save a user the hassle of selecting an answer system when a plurality of answer systems responding to a user's voice input are available), there is an obstacle to applying D2 technology to a single interactive processing system such as the D1 invention."

[4] The patentee insists, "Combining the D1 invention applied to the D2 technology with the well-known matter exemplified in D3 means a two-step change so that we cannot accept the combination. Moreover, it is premised that the D2 technology operates the mouth of the agent (simplified characters such as animals) in response to a statement, and the well-known matter exemplified in D3 is displayed as a still image, so that it is hindsight to apply the well-known matter to the D2 technology."