US 20250266034A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2025/0266034 A1**

NISHIZAWA et al. (43) **Pub. Date: Aug. 21, 2025**

(54) **SYSTEM, SUPPORT METHOD, SERVER APPARATUS AND COMMUNICATION PROGRAM**

(71) Applicants: **DAIKIN INDUSTRIES, LTD.**, Osaka (JP); **Fairy Devices Inc.**, Tokyo (JP)

(72) Inventors: **Miyu NISHIZAWA**, Osaka (JP); **Masato FUJINO**, Tokyo (JP)

(21) Appl. No.: **19/105,244**

(22) PCT Filed: **Aug. 29, 2023**

(86) PCT No.: **PCT/JP2023/031135**

§ 371 (c)(1),
(2) Date: **Feb. 20, 2025**

(30) **Foreign Application Priority Data**

Sep. 5, 2022 (JP) 2022-140702

Publication Classification

(51) **Int. Cl.**
G10L 15/01 (2013.01)
G10L 15/26 (2006.01)

(52) **U.S. Cl.**
CPC ***G10L 15/01*** (2013.01); ***G10L 15/26*** (2013.01)

(57) **ABSTRACT**

A system for converting speech data into text data includes one or more control units that determine whether or not the speech data being acquired can be converted into the text data, or whether or not a quality of the converted text data after conversion of the speech data being acquired deteriorates, and output a determination result.

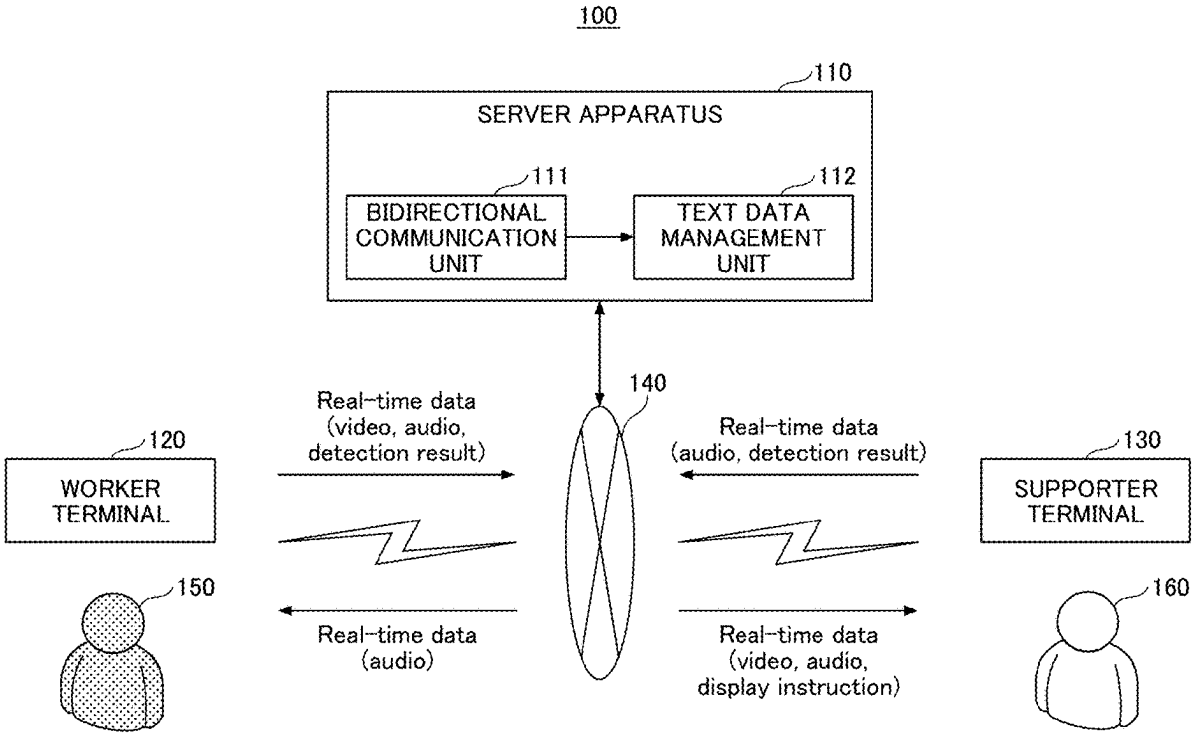


FIG.1

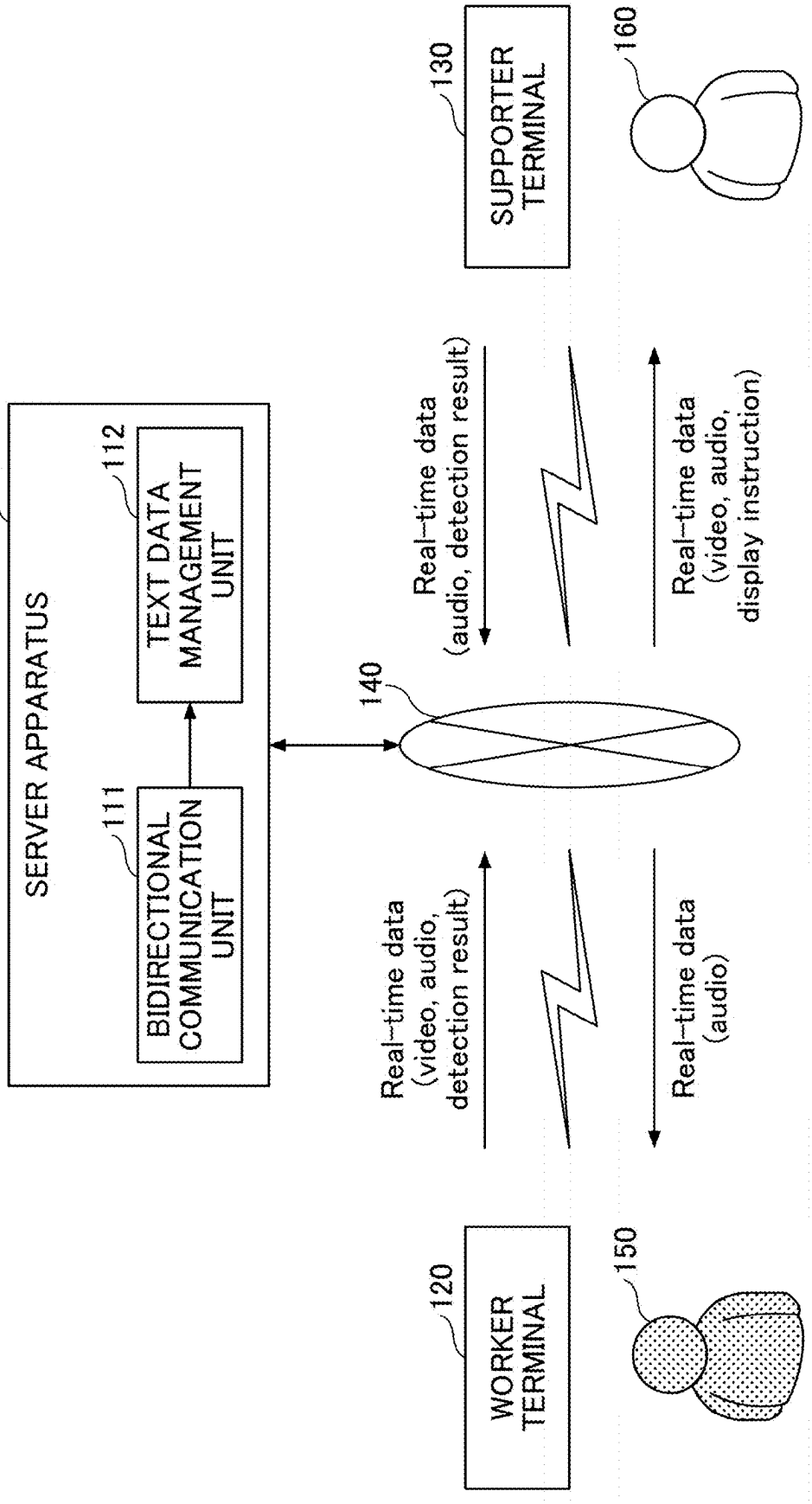
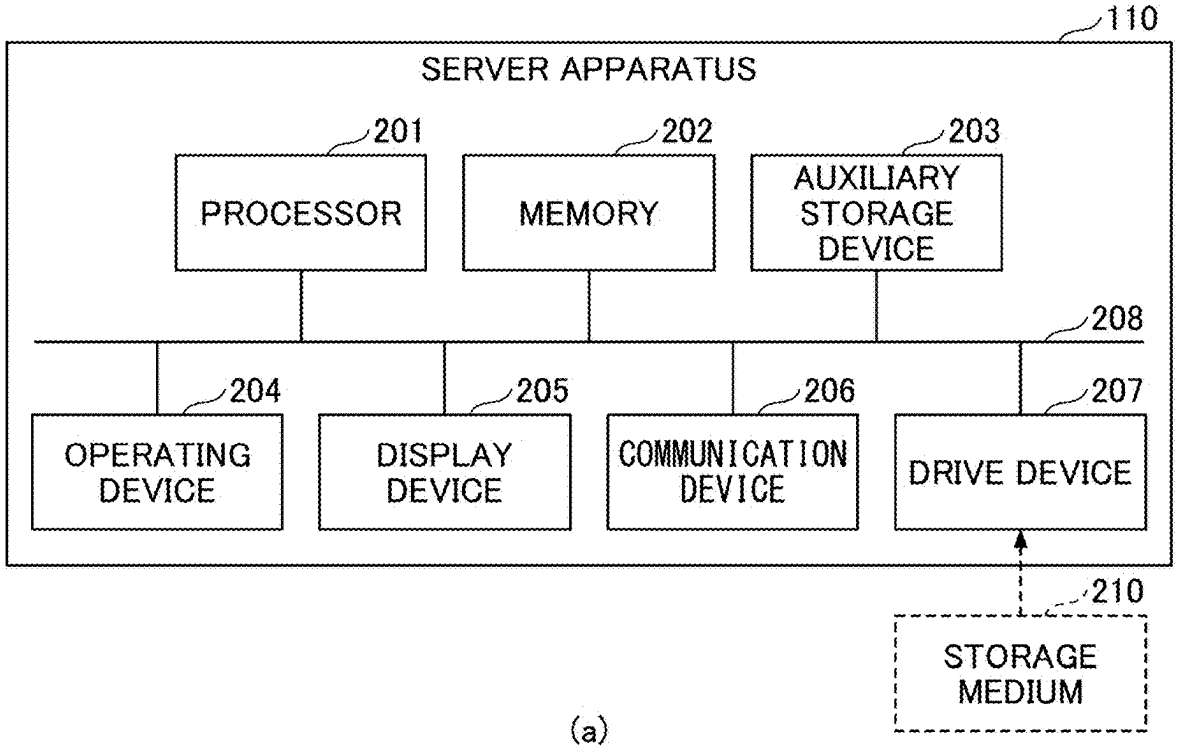
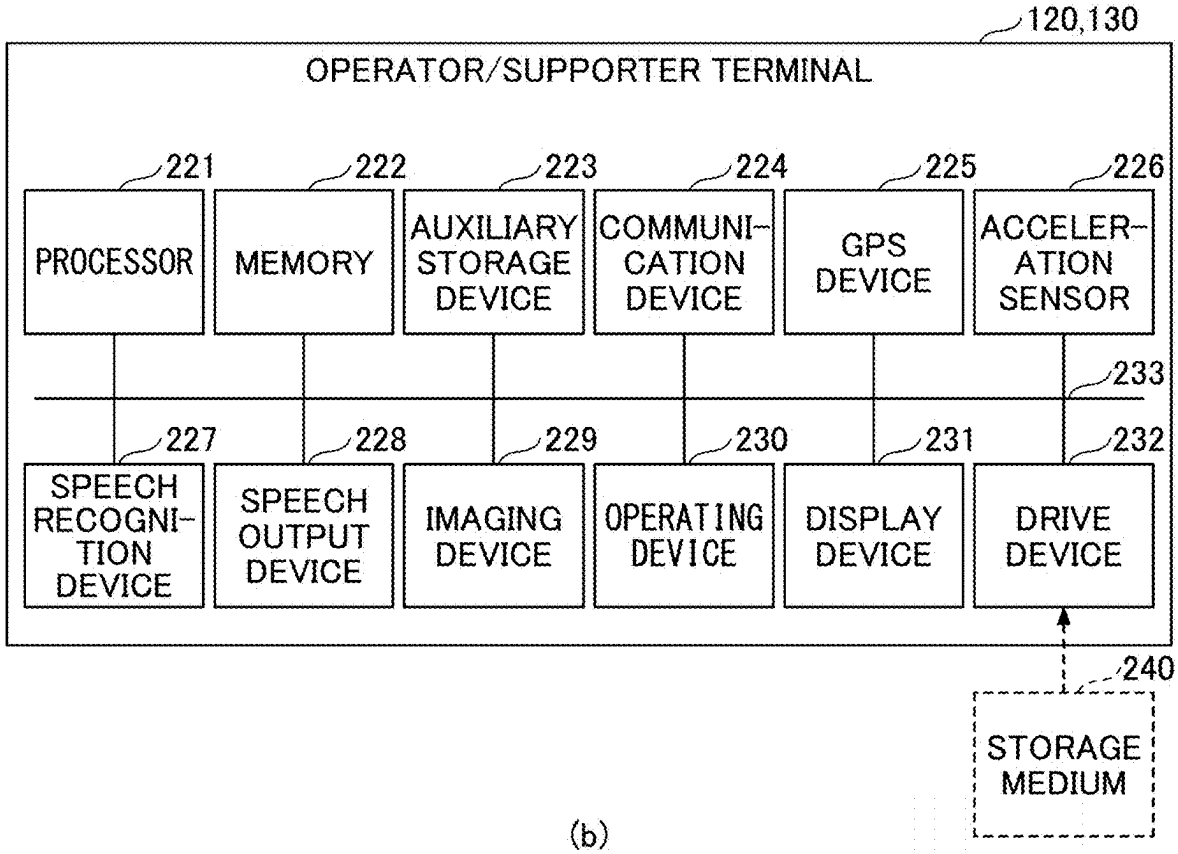

100
110
SERVER APPARATUS
111
BIDIRECTIONAL COMMUNICATION UNIT
112
TEXT DATA MANAGEMENT UNIT
140
120
WORKER TERMINAL
150
130
SUPPORTER TERMINAL
160
Real-time data (video, audio, detection result)
Real-time data (audio)
Real-time data (audio, detection result)
Real-time data (video, audio, display instruction)

FIG.2


110
SERVER APPARATUS
201
PROCESSOR
202
MEMORY
203
AUXILIARY STORAGE DEVICE
208
204
OPERATING DEVICE
205
DISPLAY DEVICE
206
COMMUNICATION DEVICE
207
DRIVE DEVICE
210
STORAGE MEDIUM

(a)


120,130
OPERATOR/SUPPORTER TERMINAL
221
PROCESSOR
222
MEMORY
223
AUXILIARY STORAGE DEVICE
224
COMMUNI-CATION DEVICE
225
GPS DEVICE
226
ACCELER-ATION SENSOR
233
227
SPEECH RECOGNI-TION DEVICE
228
SPEECH OUTPUT DEVICE
229
IMAGING DEVICE
230
OPERATING DEVICE
231
DISPLAY DEVICE
232
DRIVE DEVICE
240
STORAGE MEDIUM

(b)

FIG.3

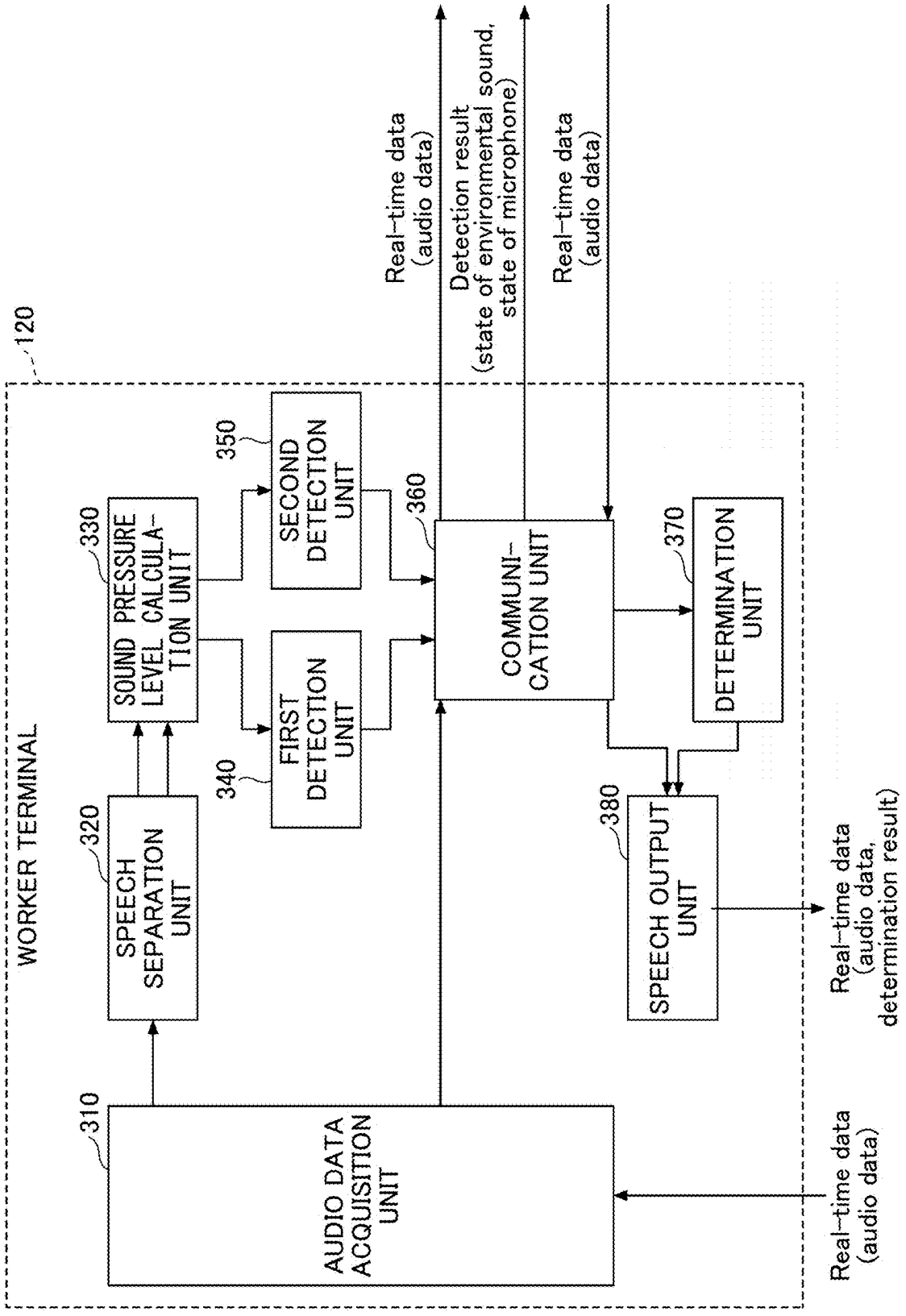
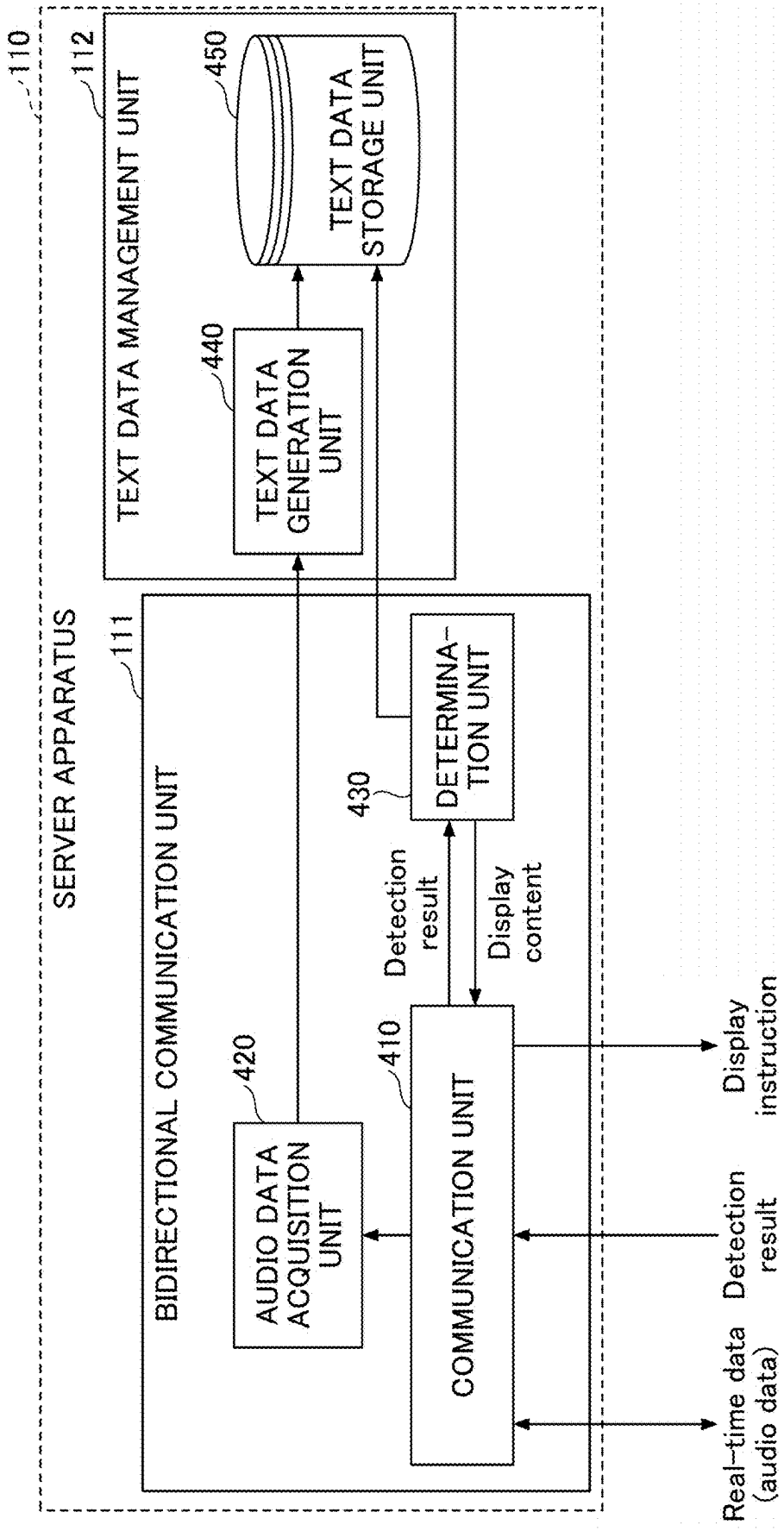

120
WORKER TERMINAL
310
AUDIO DATA ACQUISITION UNIT
320
SPEECH SEPARATION UNIT
330
SOUND PRESSURE LEVEL CALCULATION UNIT
340
FIRST DETECTION UNIT
350
SECOND DETECTION UNIT
360
COMMUNICATION UNIT
370
DETERMINATION UNIT
380
SPEECH OUTPUT UNIT
Real-time data (audio data)
Detection result (state of environmental sound, state of microphone)
Real-time data (audio data)
Real-time data (audio data)
Real-time data (audio data, determination result)

FIG.4


110
SERVER APPARATUS
112
TEXT DATA MANAGEMENT UNIT
450
TEXT DATA STORAGE UNIT
440
TEXT DATA GENERATION UNIT
111
BIDIRECTIONAL COMMUNICATION UNIT
430
DETERMINATION UNIT
Detection result
Display content
420
AUDIO DATA ACQUISITION UNIT
410
COMMUNICATION UNIT
Display instruction
Real-time data (audio data)
Detection result

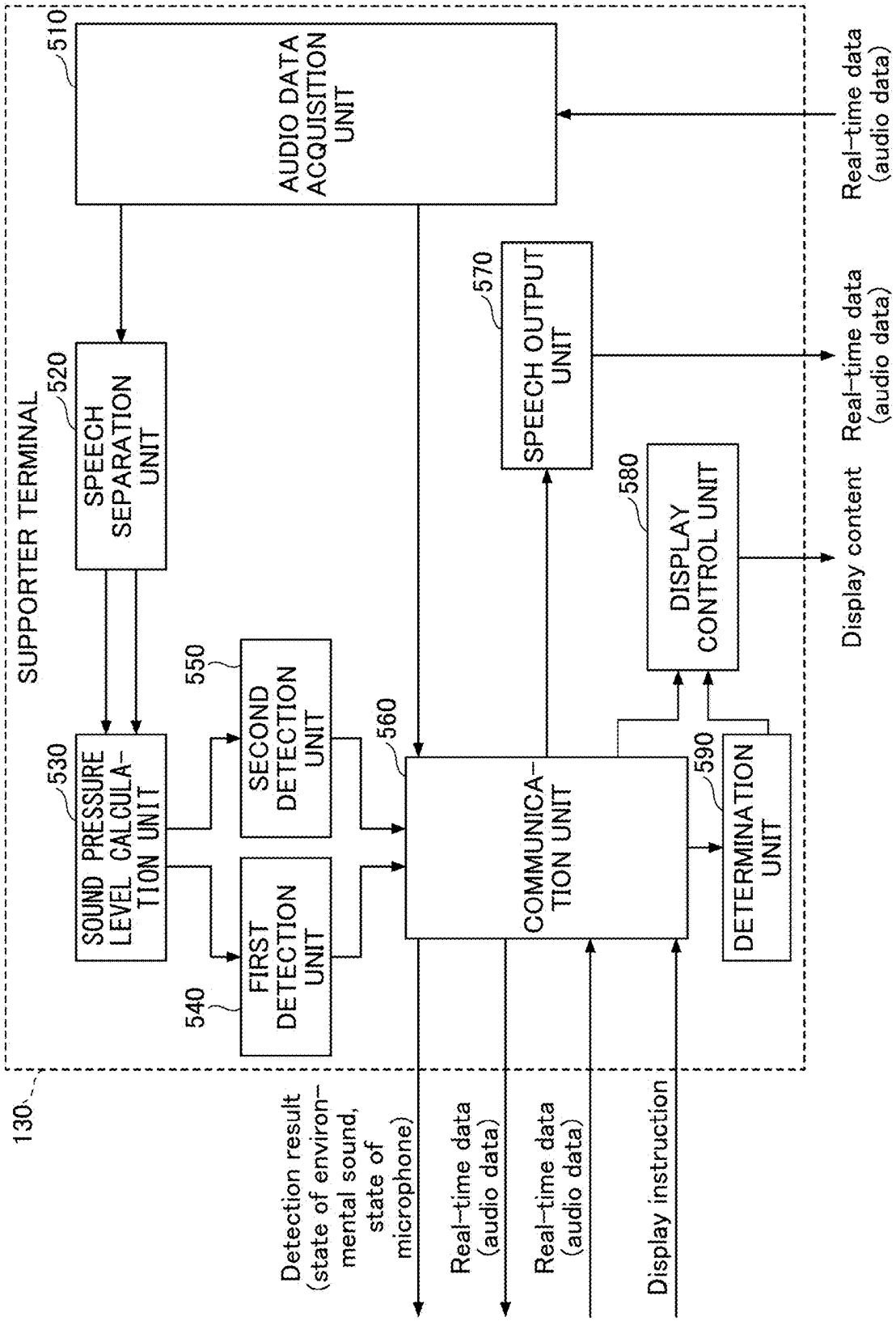

130
SUPPORTER TERMINAL
510
AUDIO DATA ACQUISITION UNIT
520
SPEECH SEPARATION UNIT
530
SOUND PRESSURE LEVEL CALCULATION UNIT
540
FIRST DETECTION UNIT
550
SECOND DETECTION UNIT
560
COMMUNICATION UNIT
570
SPEECH OUTPUT UNIT
580
DISPLAY CONTROL UNIT
590
DETERMINATION UNIT
Real-time data (audio data)
Real-time data (audio data)
Display content
Detection result (state of environmental sound, state of microphone)
Real-time data (audio data)
Real-time data (audio data)
Display instruction

FIG.5

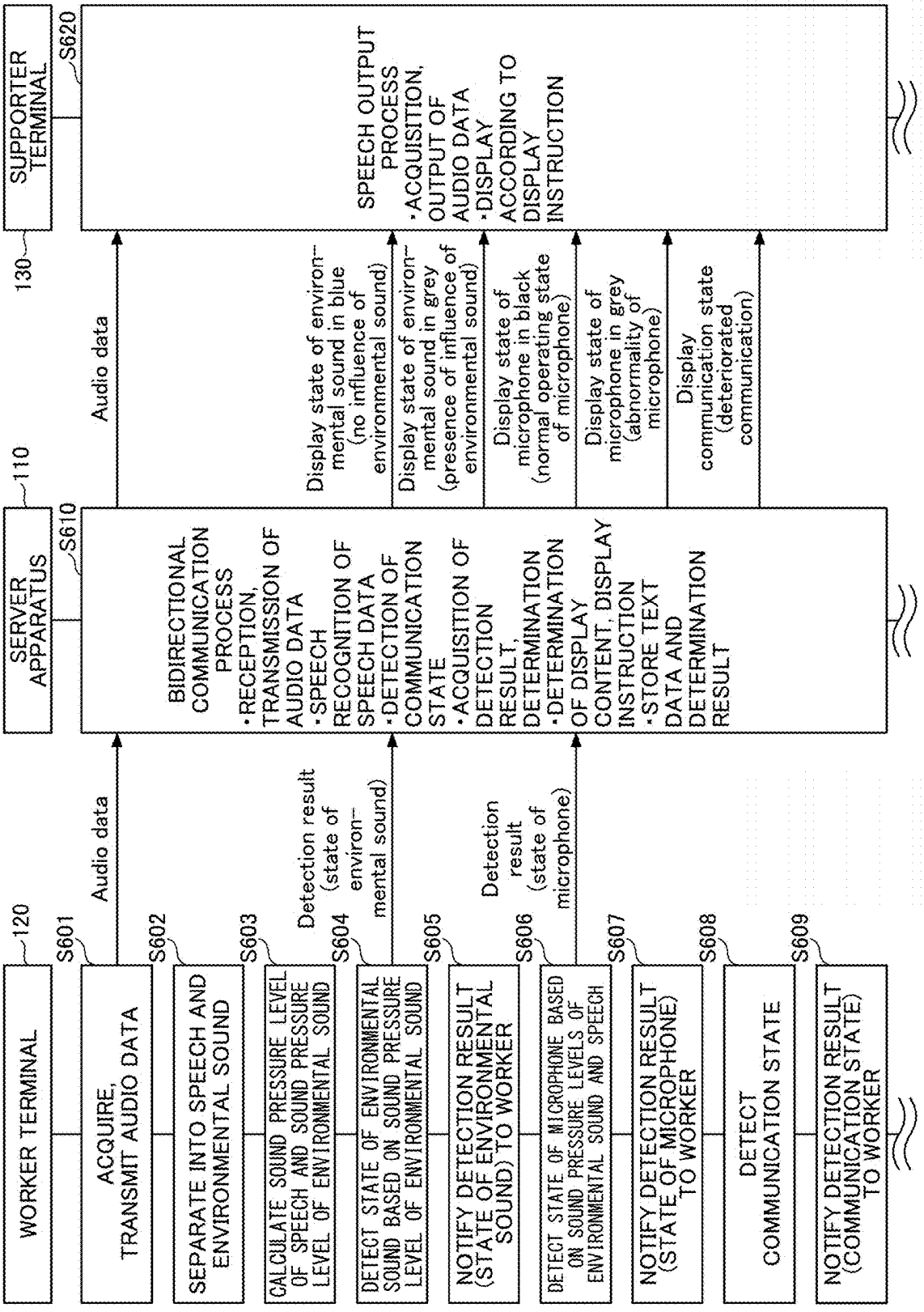

WORKER TERMINAL
120
SERVER APPARATUS
110
SUPPORTER TERMINAL
130
S601
ACQUIRE, TRANSMIT AUDIO DATA
Audio data
S610
BIDIRECTIONAL COMMUNICATION PROCESS
・RECEPTION, TRANSMISSION OF AUDIO DATA
・SPEECH RECOGNITION OF SPEECH DATA
・DETECTION OF COMMUNICATION STATE
・ACQUISITION OF DETECTION RESULT, DETERMINATION
・DETERMINATION OF DISPLAY CONTENT, DISPLAY INSTRUCTION
・STORE TEXT DATA AND DETERMINATION RESULT
Audio data
S620
SPEECH OUTPUT PROCESS
・ACQUISITION, OUTPUT OF AUDIO DATA
・DISPLAY ACCORDING TO DISPLAY INSTRUCTION
S602
SEPARATE INTO SPEECH AND ENVIRONMENTAL SOUND
S603
CALCULATE SOUND PRESSURE LEVEL OF SPEECH AND SOUND PRESSURE LEVEL OF ENVIRONMENTAL SOUND
Detection result (state of environmental sound)
Display state of environmental sound in blue (no influence of environmental sound)
S604
DETECT STATE OF ENVIRONMENTAL SOUND BASED ON SOUND PRESSURE LEVEL OF ENVIRONMENTAL SOUND
Display state of environmental sound in grey (presence of influence of environmental sound)
S605
NOTIFY DETECTION RESULT (STATE OF ENVIRONMENTAL SOUND) TO WORKER
Detection result (state of microphone)
Display state of microphone in black (normal operating state of microphone)
S606
DETECT STATE OF MICROPHONE BASED ON SOUND PRESSURE LEVELS OF ENVIRONMENTAL SOUND AND SPEECH
Display state of microphone in grey (abnormality of microphone)
S607
NOTIFY DETECTION RESULT (STATE OF MICROPHONE) TO WORKER
Display communication state (deteriorated communication)
S608
DETECT COMMUNICATION STATE
S609
NOTIFY DETECTION RESULT (COMMUNICATION STATE) TO WORKER

FIG.6

FIG.7

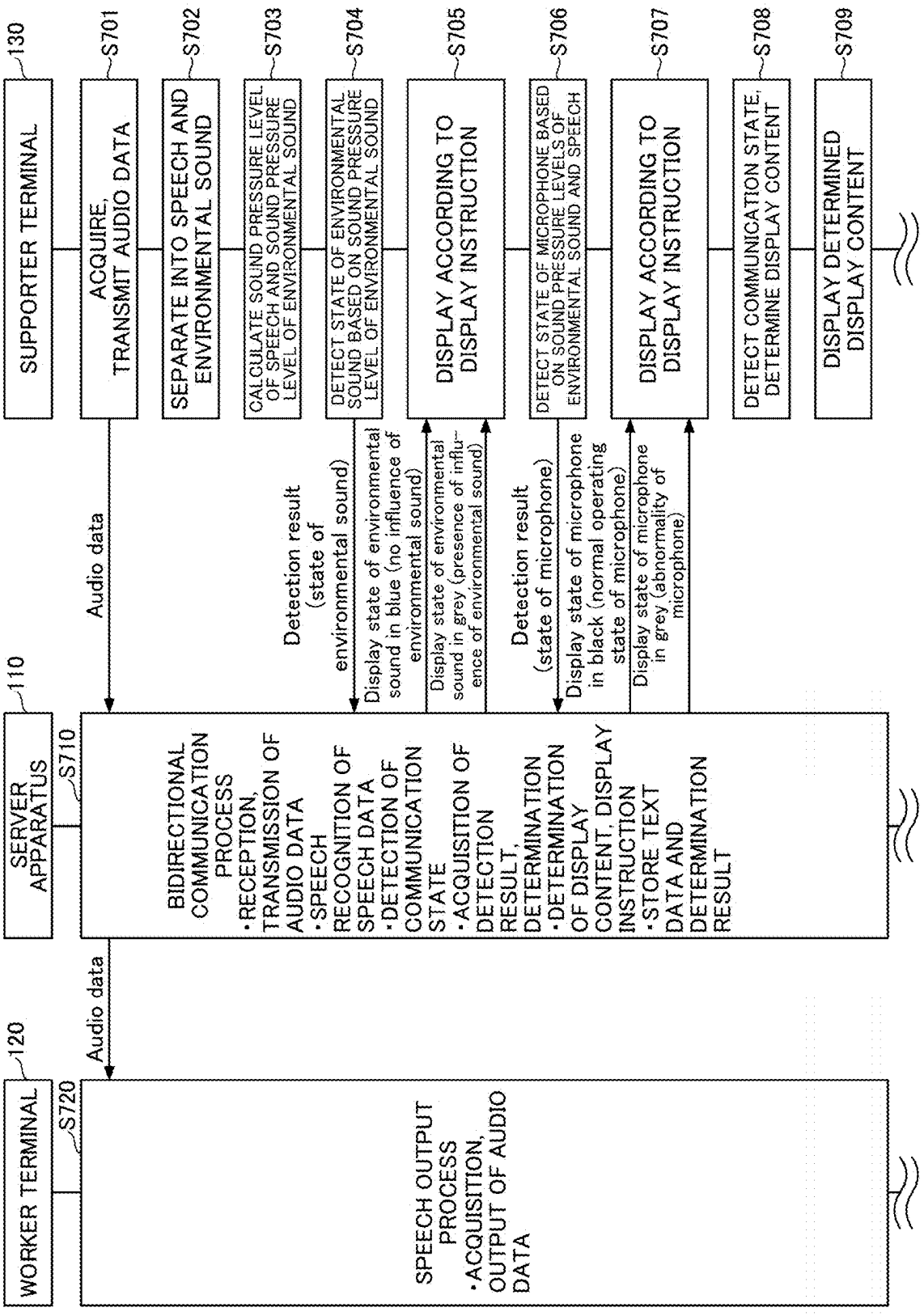

WORKER TERMINAL
120
SERVER APPARATUS
110
SUPPORTER TERMINAL
130
S720
SPEECH OUTPUT PROCESS
・ACQUISITION, OUTPUT OF AUDIO DATA
Audio data
S710
BIDIRECTIONAL COMMUNICATION PROCESS
・RECEPTION, TRANSMISSION OF AUDIO DATA
・SPEECH RECOGNITION OF SPEECH DATA
・DETECTION OF COMMUNICATION STATE
・ACQUISITION OF DETECTION RESULT, DETERMINATION
・DETERMINATION OF DISPLAY CONTENT, DISPLAY INSTRUCTION
・STORE TEXT DATA AND DETERMINATION RESULT
Audio data
S701
ACQUIRE, TRANSMIT AUDIO DATA
S702
SEPARATE INTO SPEECH AND ENVIRONMENTAL SOUND
S703
CALCULATE SOUND PRESSURE LEVEL OF SPEECH AND SOUND PRESSURE LEVEL OF ENVIRONMENTAL SOUND
S704
DETECT STATE OF ENVIRONMENTAL SOUND BASED ON SOUND PRESSURE LEVEL OF ENVIRONMENTAL SOUND
Detection result (state of environmental sound)
Display state of environmental sound in blue (no influence of environmental sound)
S705
DISPLAY ACCORDING TO DISPLAY INSTRUCTION
Display state of environmental sound in grey (presence of influence of environmental sound)
S706
DETECT STATE OF MICROPHONE BASED ON SOUND PRESSURE LEVELS OF ENVIRONMENTAL SOUND AND SPEECH
Detection result (state of microphone)
Display state of microphone in black (normal operating state of microphone)
S707
DISPLAY ACCORDING TO DISPLAY INSTRUCTION
Display state of microphone in grey (abnormality of microphone)
S708
DETECT COMMUNICATION STATE, DETERMINE DISPLAY CONTENT
S709
DISPLAY DETERMINED DISPLAY CONTENT

FIG.8

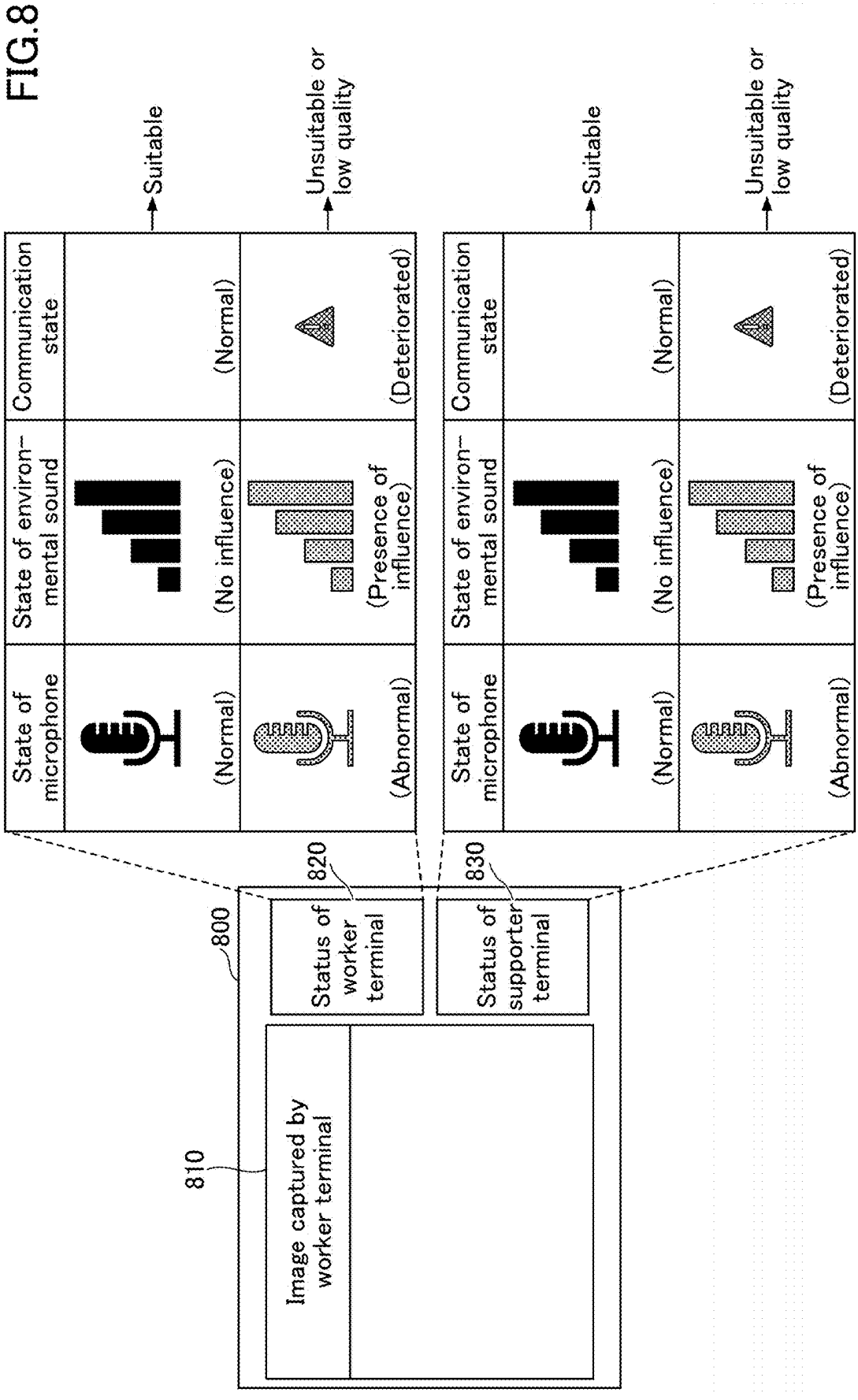

800
810
Image captured by worker terminal
820
Status of worker terminal
830
Status of supporter terminal
State of microphone
State of environmental sound
Communication state
(Normal)
(No influence)
(Normal)
Suitable
(Abnormal)
(Presence of influence)
(Deteriorated)
Unsuitable or low quality
State of microphone
State of environmental sound
Communication state
(Normal)
(No influence)
(Normal)
Suitable
(Abnormal)
(Presence of influence)
(Deteriorated)
Unsuitable or low quality

FIG.9

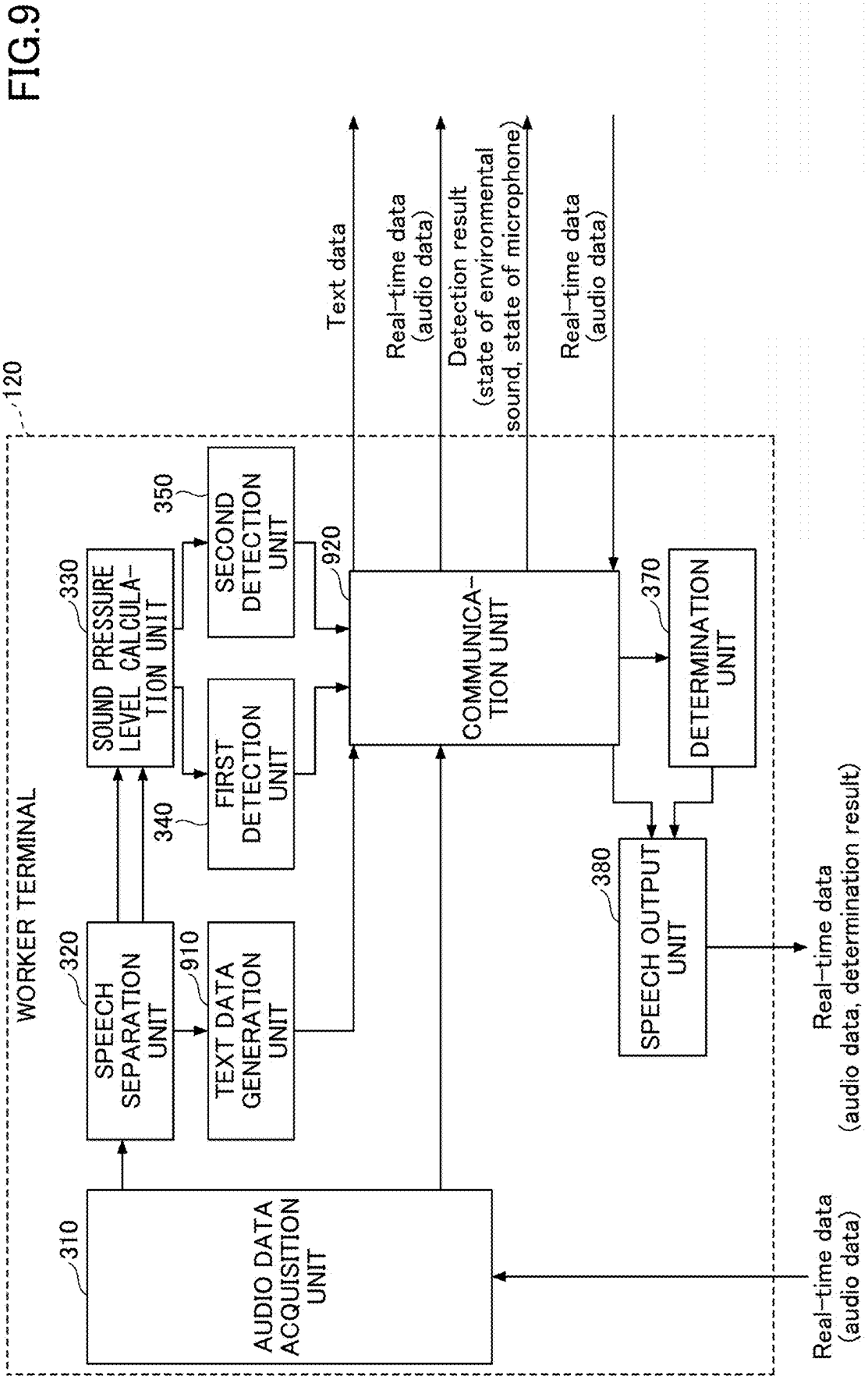
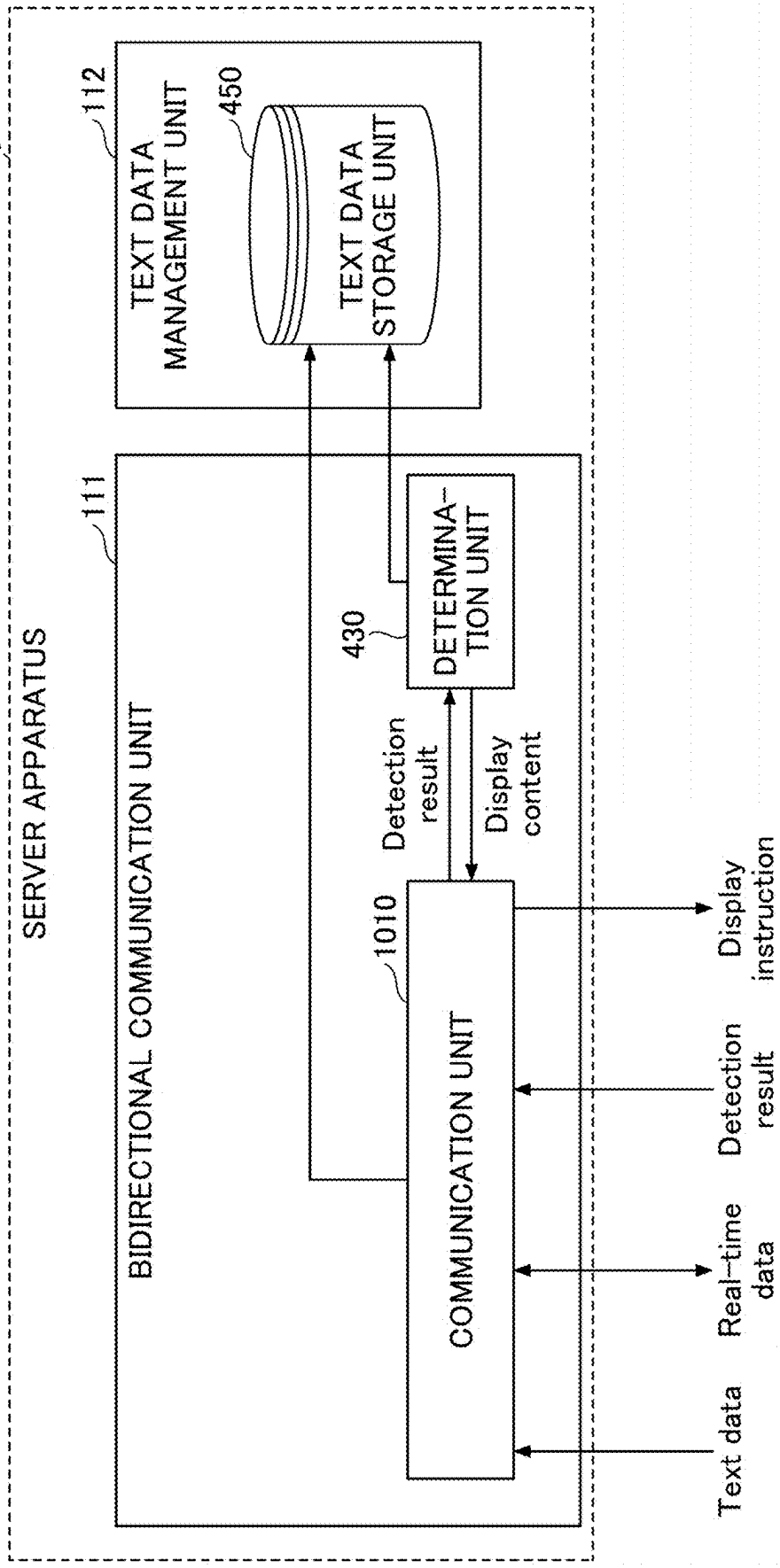

120
WORKER TERMINAL
310
AUDIO DATA ACQUISITION UNIT
320
SPEECH SEPARATION UNIT
910
TEXT DATA GENERATION UNIT
330
SOUND PRESSURE LEVEL CALCULATION UNIT
340
FIRST DETECTION UNIT
350
SECOND DETECTION UNIT
920
COMMUNICATION UNIT
370
DETERMINATION UNIT
380
SPEECH OUTPUT UNIT
Real-time data (audio data)
Text data
Real-time data (audio data)
Detection result (state of environmental sound, state of microphone)
Real-time data (audio data)
Real-time data (audio data, determination result)

FIG.10


110
SERVER APPARATUS
112
TEXT DATA MANAGEMENT UNIT
450
TEXT DATA STORAGE UNIT
111
BIDIRECTIONAL COMMUNICATION UNIT
430
DETERMINATION UNIT
Detection result
Display content
1010
COMMUNICATION UNIT
Display instruction
Detection result
Real-time data (audio data)
Text data

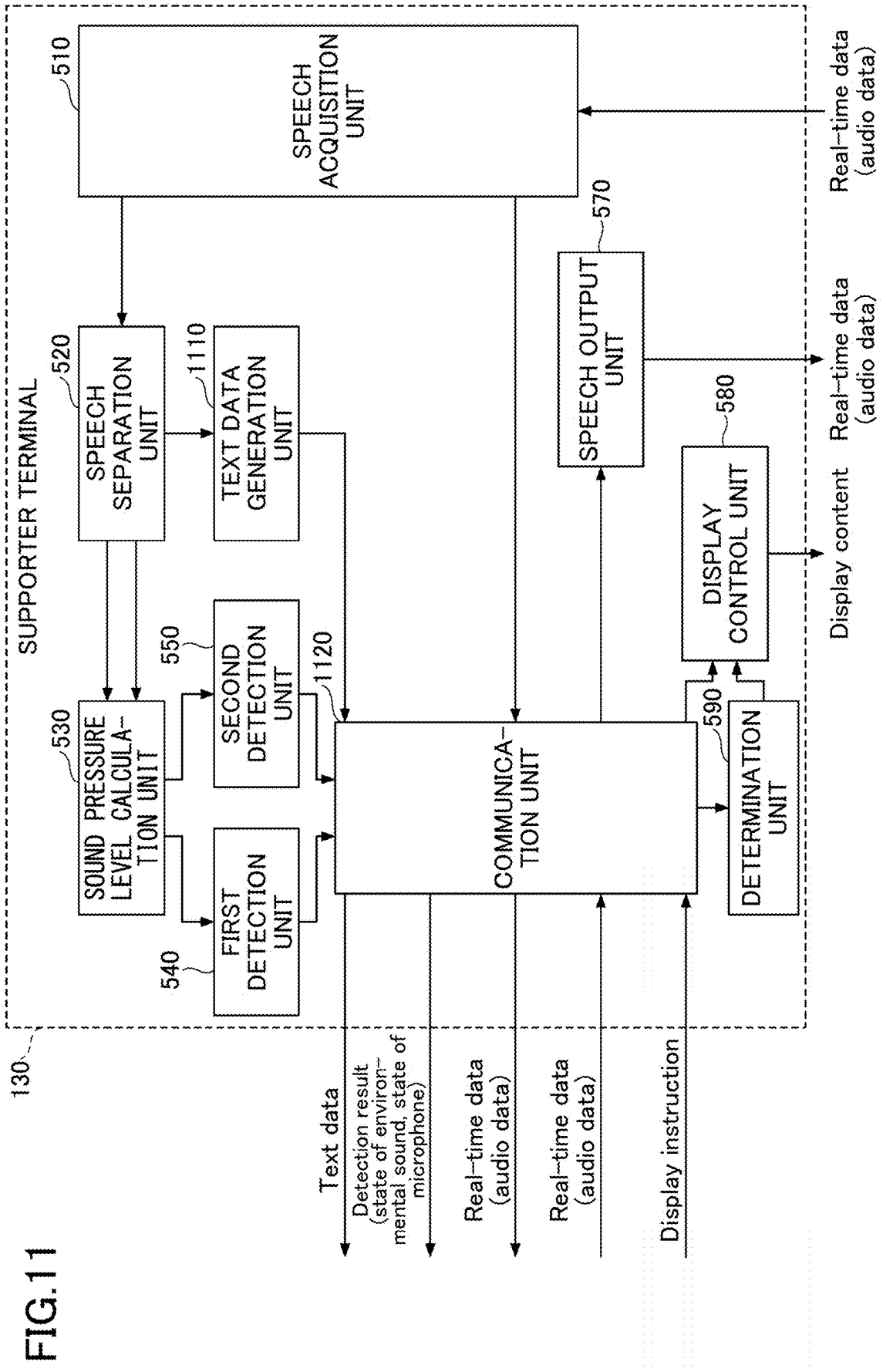

130
SUPPORTER TERMINAL
510
SPEECH ACQUISITION UNIT
520
SPEECH SEPARATION UNIT
1110
TEXT DATA GENERATION UNIT
530
SOUND PRESSURE LEVEL CALCULATION UNIT
550
SECOND DETECTION UNIT
1120
540
FIRST DETECTION UNIT
COMMUNICATION UNIT
570
SPEECH OUTPUT UNIT
580
DISPLAY CONTROL UNIT
590
DETERMINATION UNIT
Real-time data (audio data)
Real-time data (audio data)
Display content
Text data
Detection result (state of environmental sound, state of microphone)
Real-time data (audio data)
Real-time data (audio data)
Display instruction

FIG.11

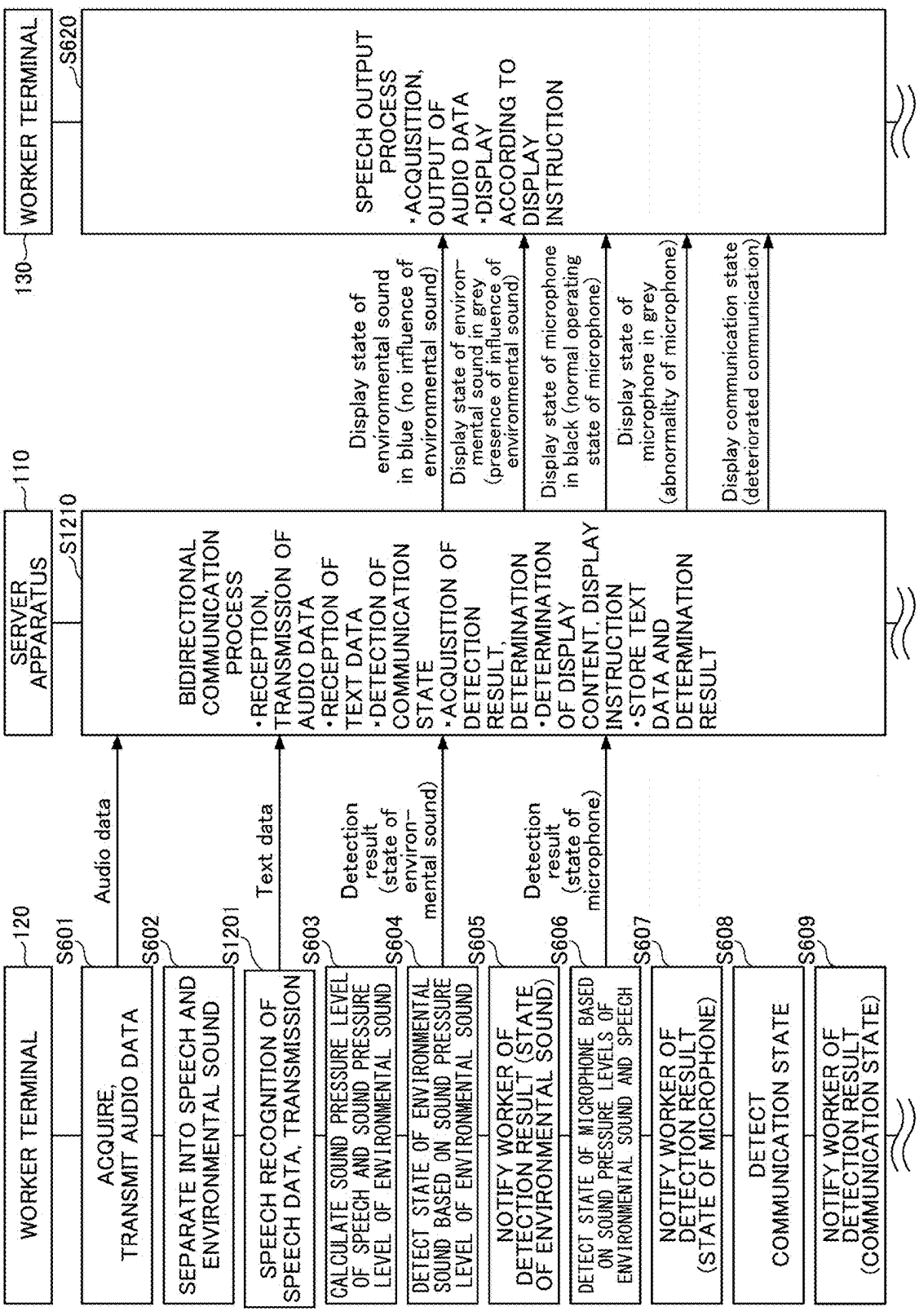

WORKER TERMINAL
120
S601
ACQUIRE, TRANSMIT AUDIO DATA
Audio data
S602
SEPARATE INTO SPEECH AND ENVIRONMENTAL SOUND
S1201
SPEECH RECOGNITION OF SPEECH DATA, TRANSMISSION
Text data
S603
CALCULATE SOUND PRESSURE LEVEL OF SPEECH AND SOUND PRESSURE LEVEL OF ENVIRONMENTAL SOUND
Detection result (state of environmental sound)
S604
DETECT STATE OF ENVIRONMENTAL SOUND BASED ON SOUND PRESSURE LEVEL OF ENVIRONMENTAL SOUND
S605
NOTIFY WORKER OF DETECTION RESULT (STATE OF ENVIRONMENTAL SOUND)
Detection result (state of microphone)
S606
DETECT STATE OF MICROPHONE BASED ON SOUND PRESSURE LEVELS OF ENVIRONMENTAL SOUND AND SPEECH
S607
NOTIFY WORKER OF DETECTION RESULT (STATE OF MICROPHONE)
S608
DETECT COMMUNICATION STATE
S609
NOTIFY WORKER OF DETECTION RESULT (COMMUNICATION STATE)
SERVER APPARATUS
110
S1210
BIDIRECTIONAL COMMUNICATION PROCESS
•RECEPTION, TRANSMISSION OF AUDIO DATA
•RECEPTION OF TEXT DATA
•DETECTION OF COMMUNICATION STATE
•ACQUISITION OF DETECTION RESULT, DETERMINATION
•DETERMINATION OF DISPLAY CONTENT, DISPLAY INSTRUCTION
•STORE TEXT DATA AND DETERMINATION RESULT
Display state of environmental sound in blue (no influence of environmental sound)
Display state of environmental sound in grey (presence of influence of environmental sound)
Display state of microphone in black (normal operating state of microphone)
Display state of microphone in grey (abnormality of microphone)
Display communication state (deteriorated communication)
130
WORKER TERMINAL
S620
SPEECH OUTPUT PROCESS
•ACQUISITION, OUTPUT OF AUDIO DATA
•DISPLAY ACCORDING TO DISPLAY INSTRUCTION

FIG.12

FIG.13

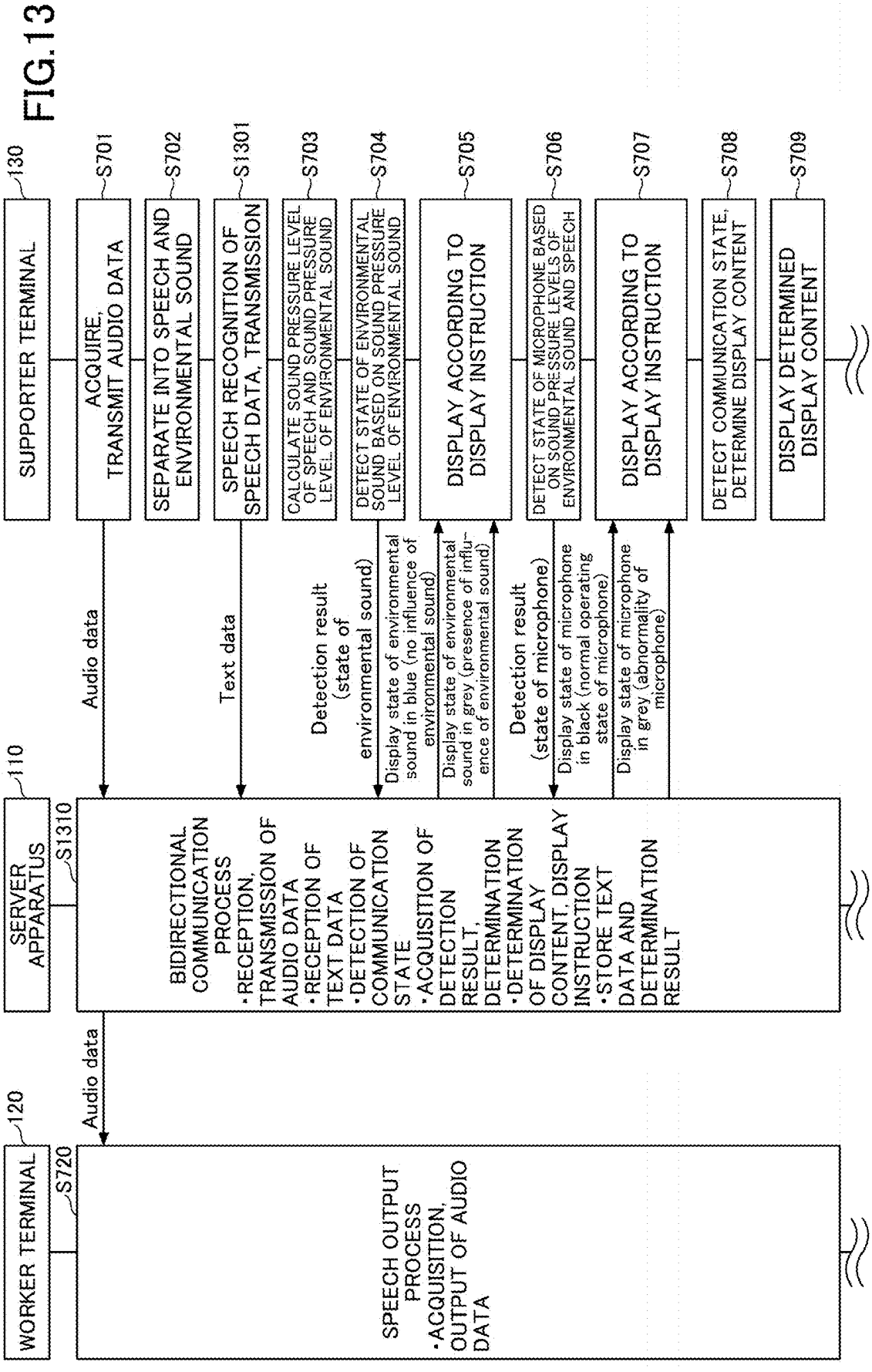

WORKER TERMINAL
120
SERVER APPARATUS
110
SUPPORTER TERMINAL
130
S720
S1310
S701
S702
S1301
S703
S704
S705
S706
S707
S708
S709
SPEECH OUTPUT PROCESS
・ACQUISITION, OUTPUT OF AUDIO DATA
BIDIRECTIONAL COMMUNICATION PROCESS
・RECEPTION, TRANSMISSION OF AUDIO DATA
・RECEPTION OF TEXT DATA
・DETECTION OF COMMUNICATION STATE
・ACQUISITION OF DETECTION RESULT, DETERMINATION
・DETERMINATION OF DISPLAY CONTENT, DISPLAY INSTRUCTION
・STORE TEXT DATA AND DETERMINATION RESULT
ACQUIRE, TRANSMIT AUDIO DATA
SEPARATE INTO SPEECH AND ENVIRONMENTAL SOUND
SPEECH RECOGNITION OF SPEECH DATA, TRANSMISSION
CALCULATE SOUND PRESSURE LEVEL OF SPEECH AND SOUND PRESSURE LEVEL OF ENVIRONMENTAL SOUND
DETECT STATE OF ENVIRONMENTAL SOUND BASED ON SOUND PRESSURE LEVEL OF ENVIRONMENTAL SOUND
DISPLAY ACCORDING TO DISPLAY INSTRUCTION
DETECT STATE OF MICROPHONE BASED ON SOUND PRESSURE LEVELS OF ENVIRONMENTAL SOUND AND SPEECH
DISPLAY ACCORDING TO DISPLAY INSTRUCTION
DETECT COMMUNICATION STATE, DETERMINE DISPLAY CONTENT
DISPLAY DETERMINED DISPLAY CONTENT
Audio data
Audio data
Text data
Detection result (state of environmental sound)
Display state of environmental sound in blue (no influence of environmental sound)
Display state of environmental sound in grey (presence of influence of environmental sound)
Detection result (state of microphone)
Display state of microphone in black (normal operating state of microphone)
Display state of microphone in grey (abnormality of microphone)

FIG.14

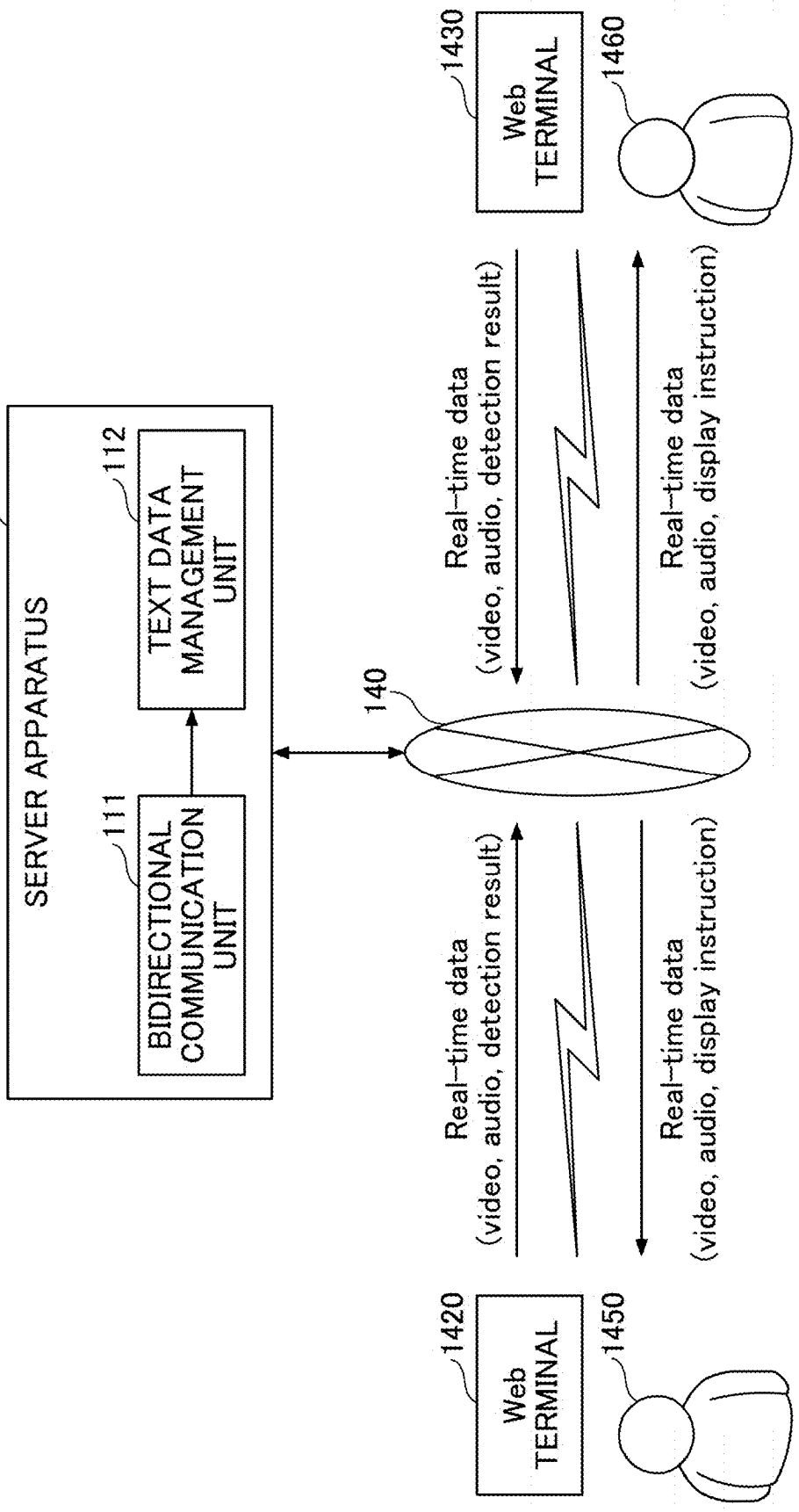

1400
110
SERVER APPARATUS
111
BIDIRECTIONAL COMMUNICATION UNIT
112
TEXT DATA MANAGEMENT UNIT
140
Real-time data (video, audio, detection result)
Real-time data (video, audio, display instruction)
Real-time data (video, audio, detection result)
Real-time data (video, audio, display instruction)
1420
Web TERMINAL
1450
1430
Web TERMINAL
1460

FIG.15

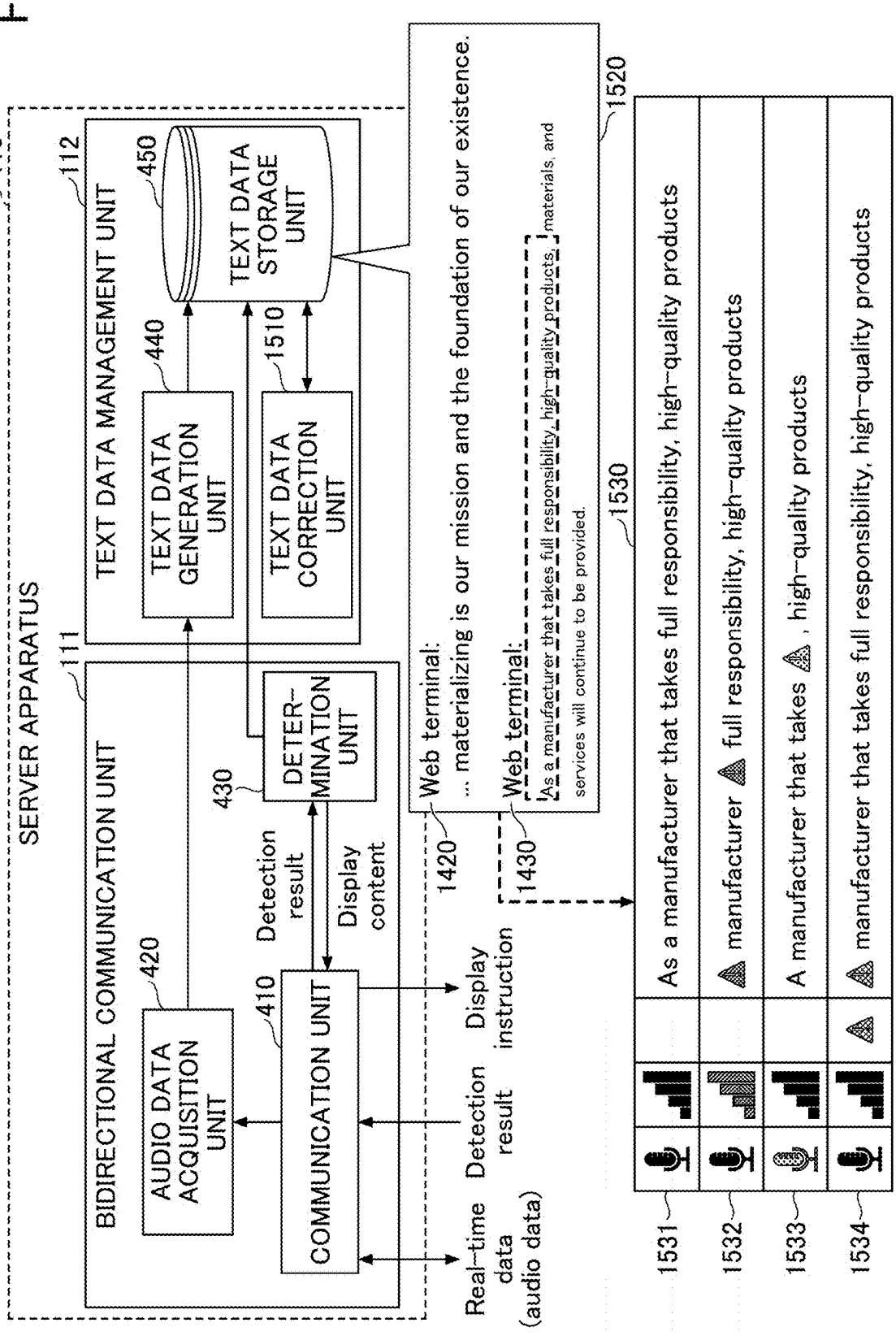
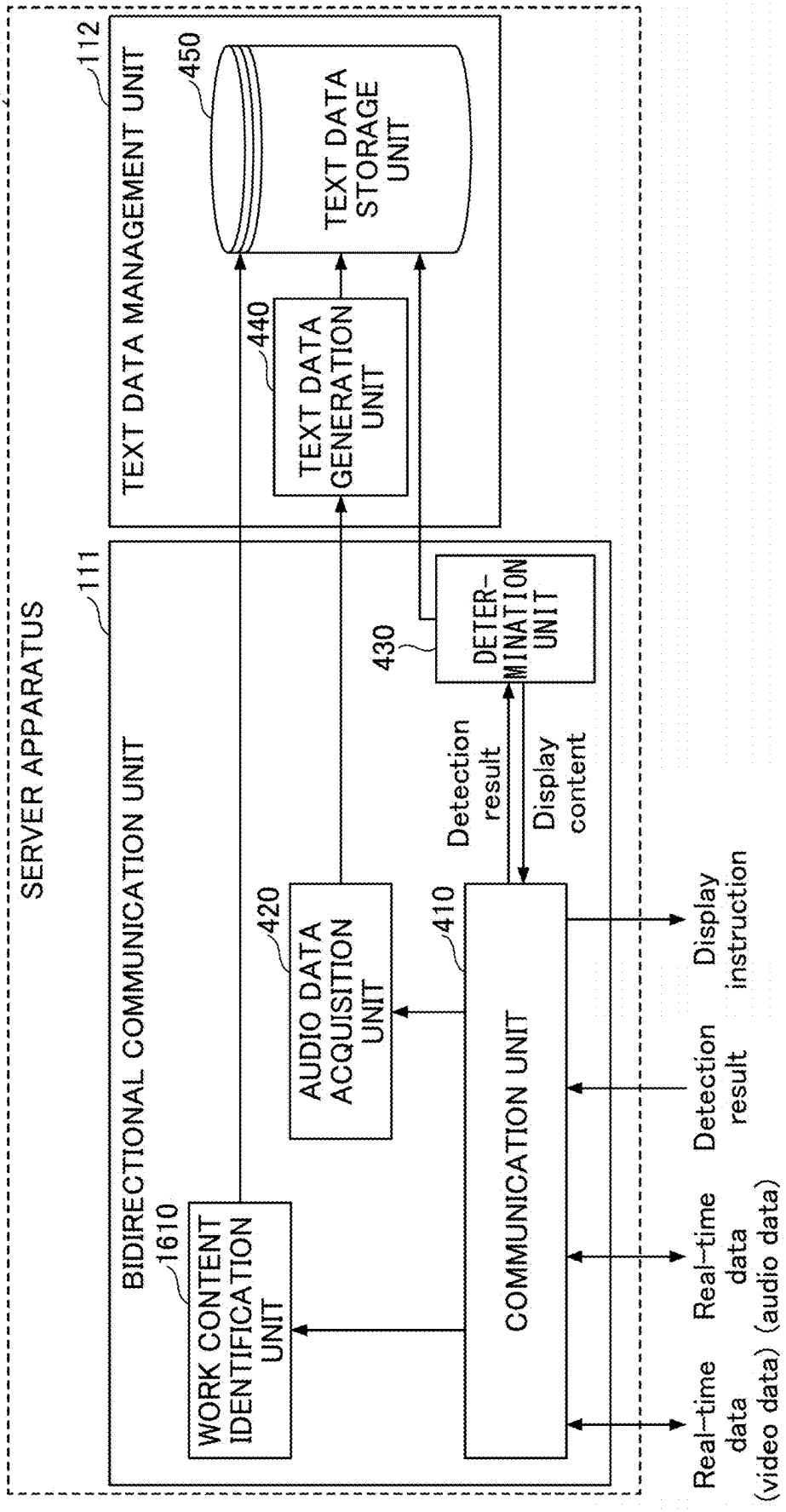

SERVER APPARATUS
110
BIDIRECTIONAL COMMUNICATION UNIT
111
AUDIO DATA ACQUISITION UNIT
420
COMMUNICATION UNIT
410
TEXT DATA MANAGEMENT UNIT
112
TEXT DATA GENERATION UNIT
440
TEXT DATA CORRECTION UNIT
1510
TEXT DATA STORAGE UNIT
450
DETER-MINATION UNIT
430
Detection result
Display content
Real-time data (audio data)
Detection result
Display instruction
Web terminal:
... materializing is our mission and the foundation of our existence.
1420
Web terminal:
As a manufacturer that takes full responsibility, high-quality products, materials, and services will continue to be provided.
1430
1520
1530
1531
As a manufacturer that takes full responsibility, high-quality products
1532
manufacturer full responsibility, high-quality products
1533
A manufacturer that takes , high-quality products
1534
manufacturer that takes full responsibility, high-quality products

FIG.16


110
SERVER APPARATUS
112
TEXT DATA MANAGEMENT UNIT
450
TEXT DATA STORAGE UNIT
440
TEXT DATA GENERATION UNIT
111
BIDIRECTIONAL COMMUNICATION UNIT
430
DETER-MINATION UNIT
Detection result
Display content
1610
WORK CONTENT IDENTIFICATION UNIT
420
AUDIO DATA ACQUISITION UNIT
410
COMMUNICATION UNIT
Display instruction
Detection result
Real-time data (audio data)
Real-time data (video data)

SYSTEM, SUPPORT METHOD, SERVER APPARATUS AND COMMUNICATION PROGRAM

TECHNICAL FIELD

[0001] The present disclosure relates to systems, support methods, server apparatuses, and communication programs.

BACKGROUND ART

[0002] A bidirectional communication system, which enables remote communication by bidirectionally transmitting and receiving video data and speech data, is known. In the bidirectional communication system, a speech-to-text function or the like may be used, for example, to enable a real-time speech recognition and conversion of a user's speech content into text data. As a result, the bidirectional communication system can manage the user's speech content as character information.

PRIOR ART DOCUMENTS

Patent Documents

[0003] Patent Document 1: Japanese Laid-Open Patent Publication No. 2021-2747

DISCLOSURE OF THE INVENTION

Problem to be Solved by the Invention

[0004] However, in the case of the bidirectional communication system described above, there are situations where a portion of the speech content cannot be recognized appropriately due to various causes, such as a malfunction of a microphone, a deterioration of a communication state, an influence of surrounding environmental sound, or the like.

[0005] On the other hand, even when it is found that a portion of the speech content is not converted into appropriate text data after the remote communication ends, it is difficult to restore the text data.

[0006] The present disclosure enables a user to identify a situation unsuitable for a speech recognition generated in a system that performs the speech recognition of a speech content.

Means of Solving the Problem

[0007] A first aspect of the present disclosure relates to a system for converting speech data into text data, comprising:

[0008] one or more control units configured to:

[0009] determine, during acquisition of the speech data, whether or not the speech data being acquired can be converted into the text data, or whether or not a quality of converted text data after conversion of the speech data being acquired deteriorates; and

[0010] output a determination result.

[0011] According to the first aspect of the present disclosure, in a system that performs a speech recognition of a speech content, a user can recognize the occurrence of a situation unsuitable for the speech recognition.

[0012] A second aspect of the present disclosure relates to the system of the first aspect, wherein the one or more control units are configured to:

[0013] indicate and output a cause the speech data cannot be converted into the text data, or a cause the quality of converted text data after the conversion deteriorates.

[0014] A third aspect of the present disclosure relates to the system of the second aspect, wherein the cause includes one cause among an influence of an environmental sound around a speech acquisition device configured to acquire the speech data, an influence of preceding speech data preceding the speech data on the speech data, and a malfunction of the speech acquisition device.

[0015] A fourth aspect of the present disclosure relates to the system of the third aspect, wherein the one or more control units are configured to:

[0016] determine that the quality of the converted text data deteriorates due to the influence of the environmental sound, in a case where a magnitude of environmental audio data acquired during the acquisition of the speech data is greater than or equal to a predetermined threshold value, or in a case where a ratio between the magnitude of the speech data and a magnitude of the environmental audio data acquired during the acquisition of the speech data satisfies a predetermined condition.

[0017] A fifth aspect of the present disclosure relates to the system of the third or fourth aspect, wherein the one or more control units are configured to:

[0018] determine that the speech data cannot be converted into the text data due to the malfunction of the speech acquisition device, in a case where both of the magnitudes of the speech data being acquired and the environmental audio data acquired during the acquisition of the speech data are less than the predetermined threshold value.

[0019] A sixth aspect of the present disclosure relates to the system of any one of the second to fifth aspects, wherein the cause includes a deterioration of a communication state of a bidirectional communication.

[0020] A seventh aspect of the present disclosure relates to the system of the sixth aspect, wherein the one or more control units are configured to:

[0021] determine that the speech data cannot be converted into the text data due to the deterioration of the communication state, in a case where the bidirectional communication is disconnected during the acquisition of the speech data.

[0022] An eighth aspect of the present disclosure relates to the system of any one of the second to seventh aspects, further comprising:

[0023] a worker terminal used by a worker; a supporter terminal used by a supporter who remotely supports the worker; and a server apparatus configured to perform a bidirectional communication between the worker terminal and the supporter terminal,

[0024] wherein a control unit included in the server apparatus converts the speech data into the text data.

[0025] A ninth aspect of the present disclosure relates to the system of the eighth aspect, further comprising:

[0026] a worker terminal used by a worker; a supporter terminal used by a supporter who remotely supports the worker; and a server apparatus configured to perform a bidirectional communication between the worker terminal and the supporter terminal,

[0027] wherein a control unit included in the server apparatus converts the speech data into the text data.

[0028] A tenth aspect of the present disclosure relates to the system of the ninth aspect, wherein:

[0029] a control unit included in the server apparatus determines, during the acquisition of the speech data, whether or not the speech data being acquired can be converted into the text data, or whether or not the quality of the converted text data after the conversion of the speech data being acquired deteriorates, and

[0030] a control unit included in the supporter terminal outputs the determination result.

[0031] An eleventh aspect of the present disclosure relates to the system of the ninth aspect, wherein:

[0032] the worker terminal includes a speech acquisition device configured to acquire the speech data, and

[0033] the control unit of the server apparatus determines that the speech data cannot be converted into the text data due to a malfunction of the speech acquisition device, in a case where both a magnitude of the speech data being acquired and the magnitude of the environmental audio data acquired during the acquisition of the speech data are less than the predetermined threshold value.

[0034] A twelfth aspect of the present disclosure related to the system of the ninth aspect, wherein the control unit included in the server apparatus determines that the speech data cannot be converted into the text data by a control unit of the worker terminal due to a deterioration of a communication state of the bidirectional communication, in a case where the speech data is not transmitted from the worker terminal for a certain period of time and the bidirectional communication is disconnected.

[0035] A thirteenth aspect of the present disclosure relates to the system of any one of the tenth to twelfth aspects, wherein a control unit included in the supporter terminal displays the determination result using an icon indicating a cause the speech data cannot be converted into the text data, or a cause the quality of the converted text data after the conversion deteriorates.

[0036] A fourteenth aspect of the present disclosure relates to the system of the tenth or eleventh aspect, wherein a control unit included in the worker terminal is configured to:

[0037] determine, during the acquisition of the speech data, whether or not the speech data being acquired can be converted into the text data, or whether or not the quality of the converted text data after the conversion the speech data being acquired deteriorates; and

[0038] notify the determination result by at least one of sound, light, and vibration.

[0039] A fifteenth aspect of the present disclosure relates to the system of any one of the ninth to fourteenth aspects, wherein the control unit of the server apparatus stores the text data and the determination result in association with each other.

[0040] A sixteenth aspect of the present disclosure relates to the system of to the fifteenth aspect, wherein a control unit included in the supporter terminal reads the stored text data and displays the read text data in a display mode according to the determination result.

[0041] A seventeenth aspect of the present disclosure relates to the system of any one of the eighth to sixteenth aspects, wherein the control unit included in the server apparatus corrects a portion that cannot be converted into the text data or a portion where the quality deteriorates, using text data before and after the portion.

[0042] An eighteenth aspect of the present disclosure relates to the system of any one of the ninth to seventeenth aspects, wherein the control unit included in the server apparatus identifies a work content of the worker from an image captured by the worker terminal, and stores the identified work content in association with the determination result.

[0043] A nineteenth aspect of the present disclosure relates to the system of any one of the ninth to eighteenth aspects, wherein the control unit included in the worker terminal stores the acquired speech data, and in a case where a request is made from the server apparatus, transmits the speech data in a range according to the request to the server apparatus.

[0044] A twentieth aspect of the present disclosure relates to a support method for a system that converts speech data into text data, the support method comprising the steps of:

[0045] determining, during acquisition of the speech data, whether or not the speech data being acquired can be converted into the text data, or whether or not a quality of converted text data after conversion of the speech data being acquired deteriorates; and

[0046] outputting a determination result of the determining.

[0047] A twenty-first aspect of the present disclosure relates to a server apparatus comprising a control unit configured to perform a bidirectional communication between a worker terminal used by a worker and a supporter terminal used by a supporter who remotely supports the worker,

[0048] wherein the control unit is configured to:

[0049] convert speech data into text data; and

[0050] transmit an instruction according to a determination result to the worker terminal or the supporter terminal, in a case where determining whether the speech data being acquired can be converted into the text data, or whether a quality of converted text data after conversion of the speech data being acquired deteriorates, during acquisition of the speech data by the worker terminal or the supporter terminal.

[0051] A twenty-second aspect of the present disclosure relates to a program causing a control unit of a server apparatus configured to perform a bidirectional communication between a worker terminal used by a worker and a supporter terminal used by a supporter who remotely supports the worker to perform a process including the steps of:

[0052] converting speech data into text data; and

[0053] transmitting an instruction according to a determination result to the worker terminal or the supporter terminal, in a case where determining whether the speech data being acquired can be converted into the text data, or whether a quality of converted text data after conversion of the speech data being acquired deteriorates, during acquisition of the speech data by the worker terminal or the supporter terminal.

BRIEF DESCRIPTION OF THE DRAWINGS

[0054] FIG. **1** is a first diagram illustrating an example of a system configuration of a bidirectional communication system.

[0055] FIG. **2** is a diagram illustrating an example of a hardware configuration of a server apparatus and examples of hardware configurations of a worker terminal and a supporter terminal.

[0056] FIG. **3** is a first diagram illustrating an example of a functional configuration of the worker terminal.

[0057] FIG. **4** is a first diagram illustrating an example of the details of the functional configuration of the server apparatus.

[0058] FIG. **5** is a first diagram illustrating an example of a functional configuration of a supporter terminal.

[0059] FIG. **6** is a first sequence diagram illustrating a flow chart of a communication process for a case where the worker terminal transmits audio data to the supporter terminal in the bidirectional communication system.

[0060] FIG. **7** is a first sequence diagram illustrating a flow chart of a communication process for a case where the supporter terminal transmits audio data to the worker terminal in the bidirectional communication system.

[0061] FIG. **8** is a diagram illustrating an example of a display screen of the supporter terminal.

[0062] FIG. **9** is a second diagram illustrating an example of the functional configuration of the worker terminal.

[0063] FIG. **10** is a second diagram illustrating an example of details of the functional configuration of the server apparatus.

[0064] FIG. **11** is a second diagram illustrating an example of the functional configuration of the supporter terminal.

[0065] FIG. **12** is a second sequence diagram illustrating a flow chart of the communication process for a case where the worker terminal transmits the audio data to the supporter terminal in the bidirectional communication system.

[0066] FIG. **13** is a second sequence diagram illustrating a flow chart of the communication process for the case where the supporter terminal transmits the audio data to the worker terminal in the bidirectional communication system.

[0067] FIG. **14** is a second diagram illustrating an example of a system configuration of the bidirectional communication system.

[0068] FIG. **15** is a diagram illustrating an example of text data usage.

[0069] FIG. **16** is a third diagram illustrating an example of the details of the functional configuration of the server apparatus.

MODE OF CARRYING OUT THE INVENTION

[0070] Hereinafter, each of embodiments will be described with reference to the accompanying drawings. In the present specification and drawings, constituent elements having substantially the same functional configuration are designated by the same reference numerals, and a redundant description thereof will be omitted.

First Embodiment

<System Configuration of Bidirectional Communication System>

[0071] First, a system configuration of a bidirectional communication system, which is an example of a system according to a first embodiment, will be described. FIG. **1** is a first diagram illustrating an example of the system configuration of the bidirectional communication system. As illustrated in FIG. **1**, a bidirectional communication system **100** includes a server apparatus **110** and client terminals (a worker terminal **120** and a supporter terminal **130**).

[0072] The present embodiment describes a case where:

[0073] A worker **150** working at a work site; and

[0074] A supporter **160** remotely supporting the worker **150** by remotely instructing various work to the worker **150**, communicate using the bidirectional communication system **100**. For this reason, as described above, the client terminals constituting the bidirectional communication system **100** in the present embodiment include the worker terminal **120** used by the worker **150** and the supporter terminal **130** used by the supporter **160**. However, the client terminals constituting the bidirectional communication system **100** are not limited to the combination of the worker terminal **120** and the supporter terminal **130**.

[0075] The server apparatus **110** provides a cloud service (a bidirectional communication service) via a network **140**. Specifically, a bidirectional communication program is installed in the server apparatus **110**, and the server apparatus **110** functions as a bidirectional communication unit **111** and a text data management unit **112** by executing the program. Thus, the server apparatus **110** provides a bidirectional communication service with respect to the worker **150** and the supporter **160**.

[0076] Specifically, in the server apparatus **110**, the bidirectional communication unit **111** is communicably connected to the worker terminal **120** and the supporter terminal **130**, and the bidirectional communication unit **111**:

[0077] Transmits video data and audio data received from the worker terminal **120** to the supporter terminal **130**;

[0078] Transmits a display instruction based on a detection result (details will be described later) or the like detected by the worker terminal **120** to the supporter terminal **130**;

[0079] Transmits the audio data received from the supporter terminal **130** to the worker terminal **120**; and

[0080] Transmits a display instruction based on a detection result (details will be described later) detected by the supporter terminal **130**, to the supporter terminal **130**.

[0081] The bidirectional communication unit **111** determines:

[0082] Whether or not speech data based on a speech of the worker **150** being acquired by the worker terminal **120** can be converted into text data; or

[0083] Whether or not a quality of converted text data after conversion of the speech data based on the speech of the worker **150** being acquired by the worker terminal **120** deteriorates,

[0084] based on the detection result or the like detected by the worker terminal **120**, and transmits a display instruction according to the determination result to the supporter terminal **130**.

[0085] Similarly, the bidirectional communication unit **111** determines:

[0086] Whether or not speech data based on a speech of the supporter **160** being acquired by the supporter terminal **130** can be converted into text data; or

[0087] Whether or not a quality of converted text data after the conversion of the speech data based on the speech of the supporter **160** being acquired by the supporter terminal **130** deteriorates,

[0088] based on the detection result detected by the supporter terminal **130**, and transmits a display instruction according to the determination result to the supporter terminal **130**.

[0089] In the server apparatus **110**, the text data management unit **112** converts the speech data, included in the audio data transmitted and received bidirectionally between the worker terminal **120** and the supporter terminal **130**, into text data in real time, and stores the text data.

[0090] The worker terminal **120** is a wearable terminal worn by the worker **150** at the work site. In the present embodiment, the work site is in an environment in which the communication quality of the bidirectional communication easily deteriorates.

[0091] The worker terminal **120** includes a function of capturing a video of the work site, and a function of detecting audio including the speech (for example, speech based on the speech of the worker **150**). The worker terminal **120** includes a function of transmitting the video data obtained by capturing the state of the work site, and the audio data including the speech data based on the speech of the worker **150** to the supporter terminal **130**, via the network **140** and the server apparatus **110**.

[0092] In addition, the worker terminal **120** detects a "state of environmental sound", a "state of a microphone", and a "communication state" which affect:

[0093] The conversion of the speech data into the text data based on the speech of the worker **150**; or

[0094] The quality of the text data after the conversion, and reports the detection result to the worker **150**, and transmits the detection result on the "state of the environmental sound" and the "state of the microphone" to the server apparatus **110**.

[0095] Moreover, the worker terminal **120** includes a function of receiving and outputting the audio data including the speech data based on the speech of the supporter **160**.

[0096] The supporter terminal **130** is installed in an office where the supporter **160** resides, and is operated by the supporter **160**, for example.

[0097] The supporter terminal **130** includes a function of detecting audio including speech (for example, speech based on the speech of the supporter **160**). The supporter terminal **130** includes a function of transmitting the audio data including the speech data based on the speech of the supporter **160** to the worker terminal **120**, via the network **140** and the server apparatus **110**.

[0098] Further, the supporter terminal **130** detects a "state of environmental sound", a "state of a microphone", and a "communication state" which affect:

[0099] The conversion of the speech data into the text data based on the speech of the supporter **160**; or

[0100] the quality of the converted text data after the conversion,

[0101] and transmits the detection result on the "state of the environmental sound" and the "state of the microphone" to the server apparatus **110**.

[0102] In addition, the supporter terminal **130** includes a function of receiving and outputting the video data obtained by capturing the state of the work site, and the audio data including the speech data based on the speech of the worker **150**.

[0103] Moreover, the supporter terminal **130** includes a function of displaying a display screen according to a display instruction (a display instruction based on the detection result detected by the worker terminal **120**, a display instruction based on the detection result detected by the supporter terminal **130**) or the like transmitted from the server apparatus **110**.

<Hardware Configuration of Each Apparatus>

[0104] Next, a hardware configuration of each apparatus (the server apparatus **110**, the worker terminal **120**, and the supporter terminal **130**) constituting the bidirectional communication system **100** will be described. FIG. **2** is a diagram illustrating an example of a hardware configuration of the server apparatus, and examples of hardware configurations of the worker terminal and the supporter terminal.

[0105] As illustrated in (a) of FIG. **2**, the server apparatus **110** includes a processor **201**, a memory **202**, an auxiliary storage device **203**, an operating device **204**, a display device **205**, a communication device **206**, and a drive device **207**. The hardware components of the server apparatus **110** are connected to one another via a bus **208**.

[0106] The processor **201** includes various computational devices, such as a central processing unit (CPU) or the like. The processor **201** reads various programs (for example, a bidirectional communication program or the like) into the memory **202**, and executes the various programs.

[0107] The memory **202** includes a main storage device, such as a read only memory (ROM), a random access memory (RA), or the like. The processor **201** and the memory **202** form a so-called computer (also referred to as a "controller"), and various functions are implemented by the computer when the processor **201** executes the various programs read into the memory **202**.

[0108] The auxiliary storage device **203** stores various programs, and various data used when the various programs are executed by the processor **201**. A text data storage unit **450** which will be described later is implemented by the auxiliary storage device **203**.

[0109] The operating device **204** operated by an administrator of the server apparatus **110** to perform various operations. The display device **205** displays results of various processes performed by the server apparatus **110**.

[0110] The communication device **206** communicates with an external device (for example, the worker terminal **120** or the supporter terminal **130**) via the network **140**.

[0111] A storage medium **210** can be loaded into the drive device **207**. The storage medium **210** includes media configured to optically, electrically, or magnetically storing information, such as a CD-ROM, a flexible disk, a magneto-optical disk, or the like. In addition, the storage medium **210** may include a semiconductor memory or the like, configured to electrically store information, such as a ROM, a flash memory, or the like.

[0112] The various programs installed in the auxiliary storage device **203** are installed by loading the distributed storage medium **210** into the drive device **207** and reading the various programs stored in the storage medium **210** by the drive device **207**, for example. Alternatively, the various programs installed in the auxiliary storage device **203** may be installed by being downloaded from the network **140** via the communication device **206**.

[0113] On the other hand, as illustrated in (b) of FIG. **2**, the worker terminal **120** and the supporter terminal **130** include a processor **221**, a memory **222**, an auxiliary storage device **223**, and a communication device **224**. In addition, the worker terminal **120** includes a global positioning system

(GPS) device **225**, an acceleration sensor **226**, and an imaging device **229**. Moreover, the worker terminal **120** and the supporter terminal **130** include a speech acquisition device **227**, a speech output device **228**, and an operating device **230**. Further, the supporter terminal **130** includes a display device **231**, and a drive device **232**. The hardware constituting the worker terminal **120** and the hardware constituting the supporter terminal **130** are connected to each other via a bus **233**.

[0114] The processor **221** includes various computational devices. such as a central processing unit (CPU) or the like. The processor **221** reads various programs (for example, a client program or the like described below) into the memory **222** and executes the various programs.

[0115] The memory **222** includes a main storage device, such as a read only memory (ROM), a random access memory (RAM), or the like. The processor **221** and the memory **222** form a so-called computer (also referred to as a "controller"), and various functions are implemented by the computer when the processor **221** executes the various programs read into the memory **222**.

[0116] The auxiliary storage device **223** stores various programs, and various information used when the various programs are executed by the processor **221**.

[0117] The communication device **224** is a device configured to transmit and receive various data (for example, video data, audio data, or the like) to and from the server apparatus **110**. The GPS device **225** detects a position of the worker terminal **120**, and outputs position data. The acceleration sensor **226** detects a movement of the worker **150** wearing the worker terminal **120**, and outputs acceleration data.

[0118] The speech acquisition device **227** detects speeches based on speeches of the worker **150** and the supporter **160**, environmental sound around the work site, environmental sound around the office, or the like, and outputs audio data.

[0119] The speech output device **228** is a device configured to output various data received from the server apparatus **110** to the worker **150** wearing the worker terminal **120** or the supporter **160** operating the supporter terminal **130** by speech.

[0120] The imaging device **229** captures an image of the surroundings (the work site) of the worker **150** wearing the worker terminal **120**, and generates video data.

[0121] The operating device **230** receives a simple operation of the worker **150** on the worker terminal **120**, such as operating a power ON/OFF switch or the like of the worker terminal **120**. Alternatively, the operating device **230** receives various operations of the supporter **160** with respect to the supporter terminal **130**.

[0122] The display device **231** is a device configured to display the video data received from the server apparatus **110**, and a display content according to a display instruction, on the supporter **160**.

[0123] A storage medium **240** can be load into the drive device **232**. In this case, the storage medium **240** is the same as the storage medium **210** described above.

[0124] In the case of the supporter terminal **130**, the various programs installed in the auxiliary storage device **223** are installed by loading the distributed storage medium **240** into the drive device **232**, and reading the various programs by the drive device **232**, for example. Alternatively, the various programs installed in the auxiliary storage device **223** may be installed by being downloaded from the network **140** via the communication device **224**.

<Functional Configuration of Worker Terminal>

[0125] Next, a functional configuration of the worker terminal **120** will be described. FIG. **3** is a first diagram illustrating an example of the functional configuration of the worker terminal. As described above, the client program is installed in the worker terminal **120**. The worker terminal **120** functions as an audio data acquisition unit **310**, a speech separation unit **320**, a sound pressure level calculation unit **330**, a first detection unit **340**, a second detection unit **350**, a communication unit **360**, a determination unit **370**, and a speech output unit **380** as illustrated in FIG. **3**, by executing the client program. In the example illustrated in FIG. **3**, functions related to processing of the video data are omitted for the sake of convenience, and only functions related to processing of audio data are extracted and illustrated (hereinafter, the processing of the audio data will be described).

[0126] The audio data acquisition unit **310** acquires the audio data detected by the speech acquisition device **227**, and notifies the speech separation unit **320** and the communication unit **360** of the audio data.

[0127] The speech separation unit **320** separates the audio data notified from the audio data acquisition unit **310** into speech data based on the speech of the worker **150**, and environmental audio data (audio data other than the speech data) based on the environmental sound around the speech acquisition device **227**. The speech separation unit **320** notifies the sound pressure level calculation unit **330** of the separated speech data and the environmental audio data.

[0128] The sound pressure level calculation unit **330** calculates a magnitude (a sound pressure level) of the speech data notified from the speech separation unit **320**, and notifies the second detection unit **350** of the magnitude of the speech data. In addition, the sound pressure level calculation unit **330** calculates the magnitude (the sound pressure level) of the environmental audio data notified from the speech separation unit **320**, and notifies the first detection unit **340** and the second detection unit **350** of the calculated magnitude of the environmental audio data.

[0129] The first detection unit **340** detects a state of the environmental sound by determining whether or not the sound pressure level of the environmental audio data notified from the sound pressure level calculation unit **330** is greater than or equal to a predetermined threshold value. Specifically, in a case where the first detection unit **340** determines that the sound pressure level of the environmental audio data is greater than or equal to the predetermined threshold value, the first detection unit **340** detects that a presence of an influence of the environmental sound.

[0130] On the other hand, in a case where the first detection unit **340** determines that the sound pressure level of the environmental audio data is less than the predetermined threshold value, the first detection unit **340** detects that there is no influence of the environmental sound. Further, the first detection unit **340** notifies the communication unit **360** of the state of the environmental sound (the presence of the influence of the environmental sound, or no influence of the environmental sound), as a detection result.

[0131] The second detection unit **350** detects a state of the speech acquisition device **227** (a state of the microphone), by determining whether or not both the sound pressure level of the environmental audio data and the sound pressure level of the speech data notified from the sound pressure level calculation unit **330** are less than a predetermined threshold value. Specifically, the second detection unit **350** detects a

malfunction of the speech acquisition device **227** (an abnormality of the microphone) in a case where the second detection unit **350** determines that both of the sound pressure levels of the environmental audio data and the speech data are less than the predetermined threshold value. On the other hand, in a case where the second detection unit **350** determines that one of the sound pressure levels of the environmental audio data and the speech data is greater than or equal to the predetermined threshold value, the second detection unit **350** detects no malfunction of the speech acquisition device **227** (a normal operating state of the microphone). Further, the second detection unit **350** notifies the communication unit **360** of the state of the microphone (the abnormality of the microphone, or the normal operating state of the microphone), as a detection result.

[0132] The communication unit **360** transmits the audio data notified from the audio data acquisition unit **310** to the server apparatus **110**, and transmits the state of the environmental sound (the presence of the influence of the environmental sound, or no influence of the environmental sound) and the state of the microphone (the abnormality of the microphone, or the normal operating state of the microphone) to the server apparatus **110**, as a detection result.

[0133] In addition, the communication unit **360** detects a communication state between the worker terminal **120** and the server apparatus **110**. Specifically, the communication unit **360** detects a deterioration of the communication in a case where the communication connection between the worker terminal **120** and the server apparatus **110** is disconnected (in a case where the audio data is not transmitted from the server apparatus **110** for a certain period of time). Further, the communication unit **360** detects a normal communication in a case where the communication connection between the worker terminal **120** and the server apparatus **110** is continued (in a case where the audio data is transmitted from the server apparatus **110**).

[0134] Moreover, the communication unit **360** notifies the determination unit **370** of the state of the environmental sound (the presence of the influence of the environmental sound, or no influence of the environmental sound), the state of the microphone (the abnormality of the microphone, or the normal operating state of the microphone), and the communication state (the deteriorated communication, or the normal communication), as a detection result.

[0135] Further, the communication unit **360** receives the audio data from the server apparatus **110**, and notifies the speech output unit **380** of the audio data.

[0136] The determination unit **370** determines:

[0137] Whether or not the speech data based on the speech of the worker **150** being acquired by the worker terminal **120** can be converted into text data; or

[0138] Whether or not a quality of the converted text data obtained by converting the speech data based on the speech of the worker **150** being acquired by the worker terminal **120** deteriorates,

[0139] based on the detection result notified from the communication unit **360**.

[0140] Specifically, in a case where the determination unit **370** is notified of the presence of the influence of the environmental sound, the determination unit **370** determines that the speech data based on the speech of the worker **150** cannot be converted into the text data, or that the quality of the converted text data deteriorates, due to the influence of the environmental sound.

[0141] In addition, in the case where the determination unit **370** is notified of the abnormality of the microphone, the determination unit **370** determines that the speech data based on the speech of the worker **150** cannot be converted into the text data, or that the quality of the converted text data deteriorates, due to the malfunction of the speech acquisition device **227**.

[0142] Moreover, in the case where the determination unit **370** is notified of the deterioration of the communication, the determination unit **370** determines that the speech data based on the speech of the worker **150** cannot be converted into the text data, or that the quality of the converted text data deteriorates, due to the deterioration of the communication.

[0143] In addition, in the case where the determination unit **370** is notified of no influence of the environmental sound, the normal operating state of the microphone, or the normal communication, the determination unit **370** determines that the speech data based on the speech of the worker **150** can be converted into the text data, or that the quality of the converted text data does not deteriorate.

[0144] The speech output unit **380** outputs the audio data notified from the communication unit **360** to the worker **150** via the speech output device **228**. Moreover, the speech output unit **380** notifies the worker **150** of the determination result notified from the determination unit **370** via the speech output device **228**. Thus, the worker **150** can recognize that a situation unsuitable for speech recognition of the speech content of the worker **150** occurred, and the cause thereof. As a result, the worker **150** can immediately take measures according to the cause, thus making it possible to avoid a situation in which it is found later that a portion of the speech content is not appropriately converted into the text data.

<Details of Functional Configuration of Server Apparatus>

[0145] Next, a functional configuration of the server apparatus **110** will be described in detail. As described above, the server apparatus **110** has the bidirectional communication program installed therein, and functions as the bidirectional communication unit **111** and the text data management unit **112** by executing the bidirectional communication program.

[0146] FIG. **4** is a first diagram illustrating an example of the details of the functional configuration of the server apparatus. As illustrated in FIG. **4**, the bidirectional communication unit **111** further includes a communication unit **410**, a speech data acquisition unit **420**, and a determination unit **430**. In addition, the text data management unit **112** further includes a text data generation unit **440**.

[0147] The communication unit **410** receives the audio data transmitted from the worker terminal **120**, transmits the audio data to the supporter terminal **130**, and notifies the speech data acquisition unit **420** of the audio data. In addition, the communication unit **410** receives the audio data transmitted from the supporter terminal **130**, transmits the audio data to the worker terminal **120**, and notifies the speech data acquisition unit **420** of the audio data.

[0148] The communication unit **410** receives the detection result (the state of the environmental sound and the state of the microphone) transmitted from the worker terminal **120** and the detection result (the state of the environmental sound and the state of the microphone) transmitted from the supporter terminal **130**, and notifies the determination unit **430** of the detection results.

[0149] The communication unit **410** detects a communication state between the worker terminal **120** and the server apparatus **110**. Specifically, the communication unit **360** detects the deterioration of the communication in a case where the communication connection between the worker terminal **120** and the server apparatus **110** is disconnected (in a case where the audio data is not transmitted from the worker terminal **120** for a certain period of time). In addition, the communication unit **360** detects a normal communication in a case where the communication connection between the worker terminal **120** and the server apparatus **110** is continued (in a case where the audio data is transmitted from the worker terminal **120**). Moreover, the communication unit **410** notifies the determination unit **430** of the communication state (the deteriorated communication, or the normal communication), as the detection result.

[0150] Similarly, the communication unit **410** detects a communication state between the supporter terminal **130** and the server apparatus **110**. Specifically, the communication unit **410** detects the deterioration of the communication in a case where the communication connection between the supporter terminal **130** and the server apparatus **110** is disconnected (in a case where the audio data is not transmitted from the supporter terminal **130** for a certain period of time). In addition, the communication unit **410** detects that the communication is normal in a case where the communication connection between the supporter terminal **130** and the server apparatus **110** is continued. Moreover, the communication unit **410** notifies the determination unit **430** of the communication state (the deteriorated communication, or the normal communication), as a detection result.

[0151] Further, the communication unit **410** transmits a display instruction including a display content (the state of the environmental sound, the state of the microphone, and the communication state) notified from the determination unit **430** to the supporter terminal **130**, in response to the notification of the detection result (the state of the environmental sound, the state of the microphone, and the communication state) for the worker terminal **120**. In addition, the communication unit **410** transmits a display instruction including a display content (the state of the environmental sound and the state of the microphone) notified from the determination unit **430** to the supporter terminal **130**, in response to the notification of the detection result (the state of the environmental sound, the state of the microphone, and the communication state) for the supporter terminal **130**.

[0152] The speech data acquisition unit **420** acquires the speech data from the audio data notified from the communication unit **410**, and notifies the text data generation unit **440** of the text data management unit **112** of the speech data.

[0153] The determination unit **430** determines:

[0154] Whether or not the speech data based on the speech of the worker **150** being acquired by the worker terminal **120** can be converted into the text data; or

[0155] Whether or not the quality of the converted text data after the conversion of the speech data based on the speech of the worker **150** being acquired by the worker terminal **120** deteriorates,

[0156] based on the detection result (the state of the environmental sound, the state of the microphone, and the communication state) for the worker terminal **120** notified from the communication unit **410**, and determines the display content (the state of the environmental sound, the state of the microphone, and the communication state) according to the determination result.

[0157] Specifically, in the case where the determination unit **430** is notified of the presence of the influence of the environmental sound, the determination unit **430** determines that the speech data based on the speech of the worker **150** cannot be converted into the text data, or the quality of the converted text data after the conversion deteriorates, due to the influence of the environmental sound. Further, the determination unit **430** determines the display content (the state of the environmental sound) according to the determination result.

[0158] In addition, in the case where the determination unit **430** is notified of the abnormality of the microphone, the determination unit **430** determines that the speech data based on the speech of the worker **150** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to a malfunction of the speech acquisition device **227**. Moreover, the determination unit **430** determines the display content (the state of the microphone) according to the determination result.

[0159] In addition, in the case where the determination unit **430** is notified of the deterioration of the communication, the determination unit **430** determines that the speech data based on the speech of the worker **150** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the deterioration of the communication. Further, the determination unit **430** determines the display content (the communication state) according to the determination result.

[0160] Moreover, in the case where the determination unit **430** is notified of no influence of the environmental sound, the normal operating state of the microphone, and the normal communication state, the determination unit **430** determines that the speech data based on the speech of the worker **150** can be converted into the text data, or that the quality of the converted text data after the conversion does not deteriorate. Further, the determination unit **430** determines the display content (the state of the environmental sound, the state of the microphone, and the communication state) according to the determination result.

[0161] Similarly, the determination unit **430** determines:

[0162] Whether or not the speech data based on the speech of the supporter **160** being acquired by the supporter terminal **130** can be converted into text data; or

[0163] Whether or not the quality of the converted text data obtained by converting the speech data based on the speech of the supporter **160** being acquired by the supporter terminal **130** deteriorates,

[0164] based on the detection result (the state of the environmental sound, the state of the microphone, and the communication state) for the supporter terminal **130** notified from the communication unit **410**, and determines the display content (the state of the environmental sound and the state of the microphone) according to the determination result.

[0165] Specifically, in the case where the determination unit **430** is notified of the presence of the influence of the environmental sound, the determination unit **430** determines that the speech data based on the speech of the supporter **160** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the influence of the environmental sound. Moreover, the

determination unit **430** determines the display content (the state of the environmental sound) according to the determination result.

[0166] In addition, in the case where the determination unit **430** is notified of the abnormality of the microphone, the determination unit **430** determines that the speech data based on the speech of the supporter **160** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the malfunction of the speech acquisition device **227**. Further, the determination unit **430** determines the display content (the state of the microphone) according to the determination result.

[0167] Moreover, in the case where the determination unit **430** is notified of the deterioration of the communication, the determination unit **430** determines that the speech data based on the speech of the worker **150** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the deterioration of the communication.

[0168] In addition, in the case where the determination unit **430** is notified of no influence of the environmental sound, the normal operating state of the microphone, and the normal communication, the determination unit **430** determines that the speech data based on the speech of the supporter **160** can be converted into the text data, or that the quality of the converted text data after the conversion does not deteriorate. Further, the determination unit **430** determines the display content (the state of the environmental sound and the state of the microphone) according to the determination result.

[0169] Further, the determination unit **430** stores the determination result in the text data storage unit **450** in association with text data which will be described below.

[0170] The text data generation unit **440** performs a speech recognition on the speech data notified from the speech data acquisition unit **420**, and converts the speech data into text data. In addition, the text data generation unit **440** stores the converted text data in the text data storage unit **450** in association with the determination result.

<Functional Configuration of Supporter Terminal>

[0171] Next, a functional configuration of the supporter terminal **130** will be described. FIG. **5** is a first diagram illustrating an example of the functional configuration of the supporter terminal. As described above, the client program is installed in the supporter terminal **130**. The supporter terminal **130** functions as an audio data acquisition unit **510**, a speech separation unit **520**, a sound pressure level calculation unit **530**, a first detection unit **540**, a second detection unit **550**, and a communication unit **560** illustrated in FIG. **5**, by executing the client program. In addition, the supporter terminal **130** functions as a speech output unit **570**, a display control unit **580**, and a determination unit **590**.

[0172] The audio data acquisition unit **510** acquires the audio data detected by the speech acquisition device **227**, and notifies the speech separation unit **520** and the communication unit **560** of the audio data.

[0173] The speech separation unit **520** separates the audio data notified from the audio data acquisition unit **510** into the speech data based on the speech of the supporter **160** and environmental audio data (audio data other than the speech data) based on the environmental sound around the speech acquisition device **227**. In addition, the speech separation unit **520** notifies the sound pressure level calculation unit **530** of the separated speech data and environmental audio data.

[0174] The sound pressure level calculation unit **530** calculates the sound pressure level of the speech data notified from the speech separation unit **520**, and notifies the second detection unit **550** of the calculated sound pressure level. In addition, the sound pressure level calculation unit **530** calculates the sound pressure level of the environmental audio data notified from the speech separation unit **520**, and notifies the first detection unit **540** and the second detection unit **550** of the calculated sound pressure level.

[0175] The first detection unit **540** detects the state of the environmental sound, by determining whether or not the sound pressure level of the environmental audio data notified from the sound pressure level calculation unit **530** is greater than or equal to a predetermined threshold value. Specifically, in the case where the first detection unit **540** determines that the sound pressure level of the environmental audio data is greater than or equal to the predetermined threshold value, the first detection unit **540** detects the presence of the influence of the environmental sound.

[0176] On the other hand, in the case where the first detection unit **540** determines that the sound pressure level of the environmental audio data is less than the predetermined threshold value, the first detection unit **540** detects that there is no influence of the environmental sound. Further, the first detection unit **540** notifies the communication unit **560** of the state of the environmental sound (the presence of the influence of the environmental sound, or no influence of the environmental sound), as a detection result.

[0177] The second detection unit **550** detects the state of the speech acquisition device **227** (the state of the microphone), by determining whether or not both the sound pressure level of the environmental audio data and the sound pressure level of the speech data notified from the sound pressure level calculation unit **530** are less than a predetermined threshold value. Specifically, in the case where the second detection unit **550** determines that both of the sound pressure levels are less than the predetermined threshold value, the second detection unit **550** detects a malfunction of the speech acquisition device **227** (the abnormality of the microphone). On the other hand, in the case where the second detection unit **550** determines that one of the sound pressure levels is greater than or equal to the predetermined threshold value, the second detection unit **550** detects no malfunction of the speech acquisition device **227** (the microphone is normal). Further, the second detection unit **550** notifies the communication unit **560** of the state of the microphone (the abnormality of the microphone, or the normal operating state of the microphone), as a detection result.

[0178] The communication unit **560** transmits the audio data notified from the audio data acquisition unit **510** to the server apparatus **110**, and transmits the state of the environmental sound (the presence of the influence of the environmental sound, or no influence of the environmental sound) and the state of the microphone (the abnormality of the microphone, or the normal operating state of the microphone) to the server apparatus **110**, as a detection result.

[0179] In addition, the communication unit **560** detects a communication state between the supporter terminal **130** and the server apparatus **110**. Specifically, in a case where the communication connection between the supporter ter-

minal **130** and the server apparatus **110** is disconnected (in a case where the audio data is not transmitted from the server apparatus **110** for a certain period of time), the communication unit **560** detects a deterioration of the communication. In addition, in a case where the communication connection between the supporter terminal **130** and the server apparatus **110** is continued (in a case where the audio data is transmitted from the server apparatus **110**), the communication unit **560** detects a normal communication.

[0180] The communication unit **560** notifies the determination unit **590** of the communication state (the deteriorated communication, or the normal communication), as a detection result.

[0181] The communication unit **560** receives the audio data from the server apparatus **110**, and notifies the speech output unit **570** of the audio data.

[0182] Further, the communication unit **560** receives a display instruction from the server apparatus **110**, and notifies the display control unit **580** of the display instruction. The display instruction received by the communication unit **560** includes:

[0183] A display instruction corresponding to the detection result (the state of the environmental sound, the state of the microphone, and the communication state) for the worker terminal **120**; and

[0184] A display instruction corresponding to the detection result (the state of the environmental sound, and the state of the microphone) for the supporter terminal **130**.

[0185] The speech output unit **570** outputs the audio data notified from the communication unit **560** to the supporter **160**, via the speech output device **228**.

[0186] The determination unit **590** determines:

[0187] Whether or not the speech data based on the speech of the worker **150** being acquired by the worker terminal **120** can be converted into the text data; or

[0188] Whether or not the quality of the converted text data obtained by converting the speech data based on the speech of the worker **150** being acquired by the worker terminal **120** deteriorates,

[0189] based on the communication state (the deteriorated communication, or the normal communication) notified from the communication unit **410** as the detection result, and determines the display content (the communication state) according to the determination result. In addition, the determination unit **590** notifies the display control unit **580** of a display instruction including the determined display content (the communication state).

[0190] The display control unit **580** displays the display instruction notified from the communication unit **560** and the display content included in the display instruction notified from the determination unit **590** on the display device **231** of the supporter terminal **130**.

<Flow Chart of Communication Process (**1**)>

[0191] Next, a flow chart of a communication process performed by the bidirectional communication system **100** will be described for a case where the audio data including the speech data based on the speech of the worker **150** is transmitted from the worker terminal **120** to the supporter terminal **130**. FIG. **6** is a first sequence diagram illustrating the flow chart of the communication process for the case where the audio data is transmitted from the worker terminal to the supporter terminal in the bidirectional communication system.

[0192] In step S**601**, the worker terminal **120** acquires the audio data including the speech data based on the speech of the worker **150**. In addition, the worker terminal **120** transmits the acquired audio data to the server apparatus **110**.

[0193] In step S**610**, the server apparatus **110** performs a bidirectional communication process. The bidirectional communication process performed by the server apparatus **110** includes an audio data reception and transmission process, a speech recognition process on the speech data, a communication state detection process, an acquisition and determination process on the detection result, a determination and display instruction process of the display content, and a storage process of the text data and the determination result. In this example, the server apparatus **110** receives the audio data, and transmits the audio data to the supporter terminal **130**. Moreover, the server apparatus **110** performs the speech recognition process on the speech data to convert the speech data into the text data.

[0194] In step S**620**, the supporter terminal **130** performs a speech output process. The speech output process performed by the supporter terminal **130** includes a process of acquiring and outputting the audio data, and a display process in response to a display instruction. In this example, the supporter terminal **130** acquires and outputs the audio data.

[0195] In step S**602**, the worker terminal **120** separates the audio data into the speech data and the environmental sound data.

[0196] In step S**603**, the worker terminal **120** calculates the sound pressure level of the audio data and the sound pressure level of the environmental sound data.

[0197] In step S**604**, the worker terminal **120** detects the state of the environmental sound, based on the sound pressure level of the environmental sound data. Further, the worker terminal **120** transmits the detection result (the state of the environmental sound) to the server apparatus **110**.

[0198] In step S**610**, the server apparatus **110** performs a bidirectional communication process. In this example, the server apparatus **110** receives the detection result (the state of the environmental sound), performs a determination based on the received detection result (the state of the environmental sound), determines a display content (the state of the environmental sound), and transmits a display instruction according to the determined display content to the supporter terminal **130**.

[0199] In a case where the server apparatus **110** receives the detection result indicating that there is no influence of the environmental sound, the server apparatus **110** determines that the speech data based on the speech of the worker **150** can be converted into the text data, or that the quality of the converted text data after the conversion does not deteriorate. In this case, the server apparatus **110** instructs the supporter terminal **130** to display an icon indicating the state of the environmental sound in blue, for example.

[0200] In addition, in a case where the server apparatus **110** receives the detection result indicating the presence of the influence of the environmental sound, the server apparatus **110** determines that the speech data based on the speech of the worker **150** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the influence of the environ-

mental sound. In this case, the server apparatus **110** instructs the supporter terminal **130** to display an icon indicating the state of the environmental sound in gray, for example.

[0201] In step S**620**, the supporter terminal **130** performs a speech output process. In this example, the supporter terminal **130** performs a display process according to the display instruction. The supporter terminal **130** displays an icon indicating the state of the environmental sound of the worker terminal **120** in blue, for example. Alternatively, the supporter terminal **130** displays. an icon indicating the state of the environmental sound of the worker terminal **120** in gray.

[0202] In step S**605**, in a case where the detection result in step S**604** indicates "the presence of the influence of the environmental sound", the worker terminal **120** notifies the worker **150** of the determination result. Specifically, the worker terminal **120** notifies the worker **150** of a determination result indicating that the speech data based on the speech of the worker **150** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the influence of the environmental sound.

[0203] In step S**606**, the worker terminal **120** detects the state of the microphone, based on the sound pressure level of the environmental sound and the sound pressure level of the speech data. The worker terminal **120** transmits a detection result (the state of the microphone) to the server apparatus **110**.

[0204] In step S**610**, the server apparatus **110** performs a bidirectional communication process. In this example, the server apparatus **110** receives the detection result (the state of the microphone), performs a determination based on the received detection result (the state of the microphone), determines the display content (the state of the microphone), and transmits a display instruction according to the determined display content to the supporter terminal **130**.

[0205] In a case where the server apparatus **110** receives the detection result indicating the normal operating state of the microphone, the server apparatus **110** determines that the speech data based on the speech of the worker **150** can be converted into the text data, or that the quality of the converted text data after the conversion does not deteriorate. In this case, the server apparatus **110** instructs the supporter terminal **130** to display an icon indicating the state of the microphone in black, for example.

[0206] In addition, in a case where the server apparatus **110** receives the detection result indicating the abnormality of the microphone, the server apparatus **110** determines that the speech data based on the speech of the worker **150** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the malfunction of the speech acquisition device **227**. In this case, the server apparatus **110** instructs the supporter terminal **130** to display an icon indicating the state of the microphone in gray, for example.

[0207] In step S**620**, the supporter terminal **130** performs a speech output process. In this example, the supporter terminal **130** performs a display process according to the display instruction. The supporter terminal **130** displays an icon indicating the state of the microphone of the worker terminal **120** in black, for example. Alternatively, the supporter terminal **130** displays an icon indicating the state of the microphone of the worker terminal **120** in gray.

[0208] In step S**607**, in a case where the detection result in step S**606** indicates "the abnormality of the microphone", the worker terminal **120** notifies the worker **150** of the determination result. Specifically, the worker terminal **120** notifies the worker **150** of the determination result indicating that the speech data based on the speech of the worker **150** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the malfunction of the speech acquisition device **227**.

[0209] In step S**608**, the worker terminal **120** detects a communication state. In addition, in step S**609**, in a case where the detection result in step S**608** indicates "the deterioration of the communication", the worker terminal **120** notifies the worker **150** of the determination result. Specifically, the worker terminal **120** notifies the worker **150** of the determination result indicating that the speech data based on the speech of the worker **150** cannot be converted into the text data, or the quality of the converted text data after the conversion deteriorates, due to the deterioration of communication.

[0210] In step S**610**, the server apparatus **110** executes a bidirectional communication process. In this example, the server apparatus **110** detects a communication state, performs a determination based on the detection result (the communication state), determines a display content (the communication state), and transmits a display instruction according to the determined display content to the supporter terminal **130**.

[0211] In a case where the server apparatus **110** acquires the detection result indicating the deterioration of the communication, the server apparatus **110** determines that the speech data based on the speech of the worker **150** cannot be converted into the text data, or the quality of the converted text data after the conversion deteriorates, due to the deterioration of the communication. In this case, the server apparatus **110** instructs the supporter terminal **130** to display an icon indicating the communication state, for example.

[0212] In step S**620**, the supporter terminal **130** performs a speech output process. In this example, the supporter terminal **130** performs a display process according to the display instruction. For example, an icon indicating the communication state between the worker terminal **120** and the server apparatus **110** is displayed.

[0213] In step S**610**, the server apparatus **110** performs a bidirectional communication process. In this example, the server apparatus **110** stores the converted text data and the determination result in association with each other.

[0214] The communication process illustrated in FIG. **6** is repeatedly performed while the worker **150** and the supporter **160** continue the communication using the bidirectional communication system **100**.

<Flow Chart of Communication Process (**2**)>

[0215] Next, a flow chart of a communication process performed by the bidirectional communication system **100** will be described for a case where the audio data including the speech data based on the speech of the supporter **160** is transmitted from the supporter terminal **130** to the worker terminal **120**. FIG. **7** is a first sequence diagram illustrating the flow chart of the communication process for the case where the audio data is transmitted from the supporter terminal to the worker terminal in the bidirectional communication system.

[0216] In step S**701**, the supporter terminal **130** acquires the audio data including the speech data based on the speech of the supporter **160**. In addition, the supporter terminal **130** transmits the acquired audio data to the server apparatus **110**.

[0217] In step S**710**, the server apparatus **110** performs a bidirectional communication process. The bidirectional communication process of the server apparatus **110** includes an audio data reception and transmission process, a speech recognition process on the speech data, a communication state detection process, an acquisition and determination process on the detection result, a determination and display instruction process of the display content, and a storage process of the text data and the determination result. In this example, the server apparatus **110** receives the audio data, and transmits the audio data to the worker terminal **120**. Further, the server apparatus **110** performs the speech recognition process on the speech data to convert the speech data into the text data.

[0218] In step S**720**, the worker terminal **120** performs a speech output process. The speech output process performed by the worker terminal **120** includes a process of acquiring and outputting the audio data. In this example, the worker terminal **120** acquires and outputs the audio data.

[0219] In step S**702**, the supporter terminal **130** separates the audio data into the speech data and the environmental sound data.

[0220] In step S**703**, the supporter terminal **130** calculates the sound pressure level of the audio data and the sound pressure level of the environmental sound data.

[0221] In step S**704**, the supporter terminal **130** detects the state of the environmental sound, based on the sound pressure level of the environmental sound data. The supporter terminal **130** transmits a detection result (the state of the environmental sound) to the server apparatus **110**.

[0222] In step S**710**, the server apparatus **110** performs a bidirectional communication process. In this example, the server apparatus **110** receives the detection result (the state of the environmental sound), performs a determination based on the received detection result (the state of the environmental sound), determines a display content (the state of the environmental sound), and transmits a display instruction according to the determined display content to the supporter terminal **130**.

[0223] In a case where the server apparatus **110** receives the detection result indicating that there is no influence of the environmental sound, the server apparatus **110** determines that the speech data based on the speech of the supporter **160** can be converted into text data, or that the quality of the converted text data after the conversion does not deteriorate. In this case, the server apparatus **110** instructs the supporter terminal **130** to display an icon indicating the state of the environmental sound in blue, for example.

[0224] In addition, in a case where the server apparatus **110** receives the detection result indicating the presence of the influence of the environmental sound, the server apparatus **110** determines that the speech data based on the speech of the supporter **160** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the influence of the environmental sound. In this case, the server apparatus **110** instructs the supporter terminal **130** to display an icon indicating the state of the environmental sound in gray, for example.

[0225] In step S**705**, the supporter terminal **130** performs a display process according to a display instruction from the server apparatus **110**. The supporter terminal **130** displays an icon indicating the state of the environmental sound of the supporter terminal **130** in blue, for example. Alternatively, the supporter terminal **130** displays an icon indicating the state of the environmental sound of the supporter terminal **130** in gray.

[0226] In step S**706**, the supporter terminal **130** detects the state of the microphone, based on the sound pressure level of the environmental sound data and the sound pressure level of the speech data. Further, the supporter terminal **130** transmits the detection result (the state of the microphone) to the server apparatus **110**.

[0227] In step S**710**, the server apparatus **110** performs a bidirectional communication process. In this example, the server apparatus **110** receives the detection result (the state of the microphone), performs a determination based on the received detection result (the state of the microphone), determines a display content (the state of the microphone), and transmits a display instruction according to the determined display content to the supporter terminal **130**.

[0228] In a case where the server apparatus **110** receives the detection result indicating the normal operating state of the microphone, the server apparatus **110** determines that the speech data based on the speech of the supporter **160** can be converted into the text data. or that the quality of the converted text data after the conversion does not deteriorate. In this case, the server apparatus **110** instructs the supporter terminal **130** to display an icon indicating the state of the microphone in black, for example.

[0229] In addition, in a case where the server apparatus **110** receives the detection result indicating the abnormality of the microphone, the server apparatus **110** determines that the speech data based on the speech of the supporter **160** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the malfunction of the speech acquisition device **227**. In this case, the server apparatus **110** instructs the supporter terminal **130** to display an icon indicating the state of the microphone in gray, for example.

[0230] In step S**707**, the supporter terminal **130** performs a display process according to the display instruction from the server apparatus **110**. The supporter terminal **130** displays an icon indicating the state of the microphone of the supporter terminal **130** in black, for example. Alternatively, the supporter terminal **130** displays an icon indicating the state of the microphone of the supporter terminal **130** in gray.

[0231] In step S**708**, the supporter terminal **130** detects a communication state, performs a determination based on the detection result (the communication state), and determines a display content (the communication state).

[0232] In a case where the detection result indicates "the deterioration of the communication", the supporter terminal **130** determines that the speech data based on the speech of the supporter **160** cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the deterioration of the communication. In this case, the supporter terminal **130** determines an icon indicating the communication state, for example, as the display content (the communication state).

[0233] In step S**709**, the supporter terminal **130** displays the determined display content. The supporter terminal **130**

displays an icon indicating the communication state between the supporter terminal **130** and the server apparatus **110**, for example.

[0234] In step S**710**, the server apparatus **110** performs a bidirectional communication process. In this example, the server apparatus **110** detects a communication state, and performs a determination based on the detection result (the communication state). Further, the server apparatus **110** stores the converted text data and the determination result in association with each other.

[0235] The communication process illustrated in FIG. **7** is performed repeatedly while the worker **150** and the supporter **160** continue the communication using the bidirectional communication system **100**.

<Display Screen Example of Supporter Terminal>

[0236] Next, an example of a display screen displayed on the display device **231** of the supporter terminal **130** will be described. FIG. **8** is a diagram illustrating an example of a display screen of the supporter terminal. As illustrated in FIG. **8**, a display screen **800** includes an area **810** in which the video data transmitted from the worker terminal **120** is displayed, an area **820** in which a status of the worker terminal **120** is displayed, and an area **830** in which a status of the supporter terminal **130** is displayed.

[0237] Among the areas of the display screen **800**, the area **810** displays the video data captured by the imaging device **229** of the worker terminal **120** and transmitted by the communication device **224** of the worker terminal **120** during the bidirectional communication process.

[0238] In the area **820**, "state of microphone", "environmental sound state", and "communication state" are displayed as the status of the worker terminal **120**.

[0239] An icon indicating the state of the microphone of the worker terminal **120** is displayed in the "state of microphone". As illustrated in FIG. **8**, in a case where the detection result (the state of the microphone) detected by the worker terminal **120** is "normal operating state of microphone", the icon indicating the state of the microphone is displayed in black. In addition, in a case where the detection result (the state of the microphone) detected by the worker terminal **120** is "abnormality of microphone", the icon indicating the state of the microphone is displayed in gray.

[0240] An icon indicating the state of the environmental sound of the worker terminal **120** is displayed in the "state of environmental sound". As illustrated in FIG. **8**, in a case where the detection result (the state of the environmental sound) detected by the worker terminal **120** is "no influence of environmental sound", the icon indicating the state of the environmental sound is displayed in black. In addition, in a case where the detection result (the state of the environmental sound) detected by the worker terminal **120** is "presence of influence of environmental sound", the icon indicating the state of the environmental sound is displayed in blue.

[0241] An icon indicating the communication state between the worker terminal **120** and the server apparatus **110** is displayed in the "communication state". As illustrated in FIG. **8**, in a case where the detection result (the communication state) detected by the server apparatus **110** is the "normal communication", no icon is displayed. On the other hand, in a case where the detection result (the communication state) detected by the server apparatus **110** indicates "deterioration of communication", an icon is displayed.

[0242] The supporter **160** refers to the area **820** while communicating with the worker **150**, and can recognize:

[0243] Whether or not the situation of the worker terminal **120** is suitable for performing the speech recognition on the speech data based on the speech of the worker **150**; and

[0244] Whether or not the situation unsuitable for the speech recognition occurred in the worker terminal **120**.

[0245] In addition, when the supporter **160** recognizes that the situation unsuitable for the speech recognition of the speech data based on the speech of the worker **150** occurred in the worker terminal **120**, the supporter **160** can understand the cause (the malfunction of the speech acquisition device **227**, the presence of the influence of the environmental sound, or the deterioration of the communication). For this reason, the supporter **160** can prompt the worker **150** to improve the situation, and can avoid a situation in which it is found later that a portion of the speech content is not converted into appropriate text data.

[0246] Similarly, in the area **830**, the "state of microphone", the "state of environmental state", and the "communication state" are displayed as the status of the supporter terminal **130**.

[0247] An icon indicating the state of the microphone of the supporter terminal **130** is displayed in the "state of microphone". As illustrated in FIG. **8**, in a case where the detection result (the state of the microphone) detected by the supporter terminal **130** is "normal operating state of microphone", the icon indicating the state of the microphone is displayed in black. In addition, in a case where the detection result (the state of the microphone) detected by the supporter terminal **130** is "abnormality of microphone", the icon indicating the state of the microphone is displayed in gray.

[0248] An icon indicating the state of the environmental sound of the supporter terminal **130** is displayed in the "state of environmental sound". As illustrated in FIG. **8**, in a case where the detection result (the state of the environmental sound) detected by the supporter terminal **130** is "no influence of environmental sound", the icon indicating the state of the environmental sound is displayed in black. In addition, in a case where the detection result (the state of the environmental sound) detected by the supporter terminal **130** is "presence of influence of environmental sound", the icon indicating the state of the environmental sound is displayed in blue.

[0249] An icon indicating the communication state between the supporter terminal **130** and the server apparatus **110** is displayed in the "communication state". As illustrated in FIG. **8**, in a case where the detection result (the communication state) detected by the supporter terminal **130** is "normal communication", the icon is not displayed. On the other hand, in a case where the detection result (the communication state) detected by the supporter terminal **130** indicates "deteriorated communication", an icon is displayed.

[0250] The supporter **160** refers to the area **830** while communicating with the worker **150**, and can recognize:

[0251] Whether or not the supporter terminal **130** is in a situation suitable for performing the speech recognition on the speech data based on the speech of the supporter **160**; and

[0252] Whether or not a situation unsuitable for the speech recognition of the speech of the supporter **160** occurred in the supporter terminal **130**.

[0253] In addition, when the supporter **160** recognizes that a situation unsuitable for performing the speech recognition on the speech data based on the speech of the supporter **160** occurred in the supporter terminal **130**, the supporter **160** can understand the cause (the malfunction of the speech acquisition device **227**, the presence of the influence of the environmental sound, or the deterioration of the communication). For this reason, the supporter **160** can immediately improve the situation, and can avoid a situation in which it is found later that a portion of the speech content is not converted into appropriate text data.

Conclusion

[0254] As is clear from the description given heretofore, the bidirectional communication system **100**, which is an example of the system according to the first embodiment, performs the following:

[0255] During the acquisition of the audio data, sound pressure data of the audio data being acquired and sound pressure data of the environmental audio data other than the audio data are calculated. Thus, it can be determined whether or not the speech data being acquired can be converted into the text data, or whether or not the quality of the converted text data after the conversion of the speech data being acquired deteriorates.

[0256] The determination result is notified to the worker, or the display content according to the determination result is displayed on the supporter terminal.

[0257] Hence, according to the bidirectional communication system **100**, the worker or the supporter can recognize the occurrence of a situation unsuitable for the speech recognition.

Second Embodiment

[0258] In the first embodiment, the case where the process of converting the speech data into the text data is collectively performed in the server apparatus **110** is described. However, the process of converting the speech data in bulk into the text data may be performed in each client terminal of the worker terminal **120** and the supporter terminal **130**. A second embodiment will be described below, focusing on the difference from the first embodiment.

<Functional Configuration of Worker Terminal>

[0259] First, the functional configuration of the worker terminal **120** according to the second embodiment will be described. FIG. **9** is a second diagram illustrating an example of the functional configuration of the worker terminal. The difference from the functional configuration of the first embodiment described with reference to FIG. **3** is that a text data generation unit **910** is provided, and functions of a communication unit **920** are different from the functions of the communication unit **360** illustrated in FIG. **3**.

[0260] The text data generation unit **910** converts the speech data notified from the speech separation unit **320** into the text data by performing the speech recognition, and notifies the communication unit **920** of the converted text data.

[0261] The communication unit **920** has a function of transmitting the text data notified from the text data generation unit **910** to the server apparatus **110**, in addition to the functions of the communication unit **360** illustrated in FIG. **3**.

<Details of Functional Configuration of Server Apparatus>

[0262] Next, the functional configuration of the server apparatus **110** according to the second embodiment will be described in detail. FIG. **10** is a second diagram illustrating an example of the details of the functional configuration of the server apparatus. The difference from the details of the functional configuration described with reference to FIG. **4** is that the functions of a communication unit **1010** are different from the functions of the communication unit **410** illustrated in FIG. **4**, and that the speech data acquisition unit **420** and the text data generation unit **440** are not provided.

[0263] In addition to the functions of the communication unit **410** illustrated in FIG. **4**, the communication unit **1010** further includes:

[0264] A function of receiving the text data transmitted from the worker terminal **120**, and storing the received text data in the text data storage unit **450** in association with the determination result determined by the determination unit **430**; and

[0265] A function of receiving the text data transmitted from the supporter terminal **130**, and storing the received text data in the text data storage unit **450** in association with the determination result determined by the determination unit **430**.

<Functional Configuration of Supporter Terminal>

[0266] Next, the functional configuration of the supporter terminal **130** in the second embodiment will be described. FIG. **11** is a second diagram illustrating an example of the functional configuration of the supporter terminal. The difference from the functional configuration described with reference to FIG. **5** is that a text data generation unit **1110** is provided and that functions of a communication unit **1120** are different from the functions of the communication unit **560** illustrated in FIG. **5**.

[0267] The text data generation unit **1110** converts the speech data notified from the speech separation unit **520** into the text data by performing the speech recognition, and notifies the communication unit **1120** of the converted text data.

[0268] The communication unit **1120** has a function of transmitting the text data notified from the text data generation unit **1110** to the server apparatus **110**, in addition to the functions of the communication unit **560** illustrated in FIG. **3**.

<Flow Chart of Communication Process (1)>

[0269] Next, a flow chart of a communication process performed by the bidirectional communication system **100** according to the second embodiment will be described for a case where the audio data including the speech data based on the speech of the worker **150** is transmitted from the worker terminal **120** to the supporter terminal **130**. FIG. **12** is a second sequence diagram illustrating the flow chart of the communication process for the case where the audio data is transmitted from the worker terminal to the supporter terminal in the bidirectional communication system. The dif-

ference from the first sequence diagram described with reference to FIG. **6** is the provision of steps S**1201** and S**1210**.

[0270] In step S**1201**, the worker terminal **120** performs a speech recognition on the speech data, and converts the speech data into text data. The worker terminal **120** transmits the converted text data to the server apparatus **110**.

[0271] In step S**1210**, the server apparatus **110** performs a bidirectional communication process. The bidirectional communication process performed by the server apparatus **110** includes an audio data reception and transmission process, a text data reception process, a communication state detection process, an acquisition and determination process on the detection result, a determination and display instruction process of the display content, and a storage process of the text data and the determination result. In this example, the server apparatus **110** stores the received text data in association with the determination result.

<Flow Chart of Communication Process (**2**)>

[0272] Next, a flow chart of a communication process performed by the bidirectional communication system **100** according to the second embodiment will be described for a case where the audio data including the speech data based on the speech of the supporter **160** is transmitted from the supporter terminal **130** to the worker terminal **120**. FIG. **13** is a second sequence diagram illustrating the flow chart of the communication process for the case where the audio data is transmitted from the supporter terminal to the worker terminal in the bidirectional communication system. The difference from the first sequence diagram described with reference to FIG. **7** is the provision of steps S**1301** and S**1310**.

[0273] In step S**1301**, the supporter terminal **130** performs a speech recognition on the speech data, and converts the speech data into text data. The supporter terminal **130** transmits the converted text data to the server apparatus **110**.

[0274] In step S**1310**, the server apparatus **110** performs a bidirectional communication process. The bidirectional communication process performed by the server apparatus **110** includes an audio data reception and transmission process, a text data reception process, a communication state detection process, an acquisition and determination process on the detection result, a determination and display instruction process of the display content, and a storage process of the text data and the determination result. In this example, the server apparatus **110** stores the received text data in association with the determination result.

Conclusion

[0275] As is clear from the description given heretofore, the bidirectional communication system **100**, which is an example of the system according to the second embodiment, includes the same functions as those of the first embodiment. Further, the bidirectional communication system **100**, which is an example of the system according to the second embodiment, is configured to perform the speech recognition process on the speech data in each client terminal, instead of performing the speech recognition process in the server apparatus **110** in a bulk.

[0276] Hence, according to the second embodiment, it is possible to obtain the same effects as those obtainable by the first embodiment, and to avoid a situation in which an accuracy of the speech recognition deteriorates due to noise that is added when the audio data is transmitted between each client terminal and the server apparatus.

Third Embodiment

[0277] The first and second embodiments are described above for the case where the bidirectional communication system is applied to the communication between:

[0278] The worker **150** working at the work site; and

[0279] The supporter **160** who remotely supports the worker **150** by remotely giving various work instructions to the worker **150**.

However, the application of the bidirectional communication system is not limited to such a case, and the communication may be applied to other cases as long as the communication is performed remotely. A third embodiment will be described for a case where the bidirectional communication system is applied to a situation in which users at remote locations hold a conference.

<System Configuration of Bidirectional Communication System>

[0280] First, a system configuration of the bidirectional communication system, which is an example of the system according to the third embodiment, will be described. FIG. **14** is a second diagram illustrating an example of the system configuration of the bidirectional communication system. As illustrated in FIG. **14**, a bidirectional communication system **1400** includes a server apparatus **110**, and client terminals (Web terminals **1420** and **1430**).

[0281] The present embodiment will be described for a case where the bidirectional communication system is applied to a situation in which the conference is held between:

[0282] A user **1450**; and

[0283] A user **1460**.

[0284] The server apparatus **110** is the same as the server apparatus **110** described in the first or second embodiment.

[0285] In addition, the Web terminals **1420** and **1430** are the same as the supporter terminal **130** described in the first or second embodiment.

[0286] The text data and the determination result stored in the text data storage unit **450** (not illustrated in FIG. **14**) in the third embodiment may be downloaded by the user **1450** or **1460** after the conference ends, and may be used as conference minutes, for example.

<Example of Text Data Usage>

[0287] Next, an example of text data usage in the bidirectional communication system **1400**, which is an example of the system according to the third embodiment, will be described. FIG. **15** is a diagram illustrating an example of the text data usage. In the case of the example of the text data usage illustrated in FIG. **15**, the text data management unit **112** further includes a text data correction unit **1510**.

[0288] The text data correction unit **1510** corrects the text data stored in the text data storage unit **450**, based on the associated determination result. Correction functions of the text data correction unit **1510** include:

[0289] A function of visualizing the determination result by adding a mark indicating a cause of a predetermined determination result (the malfunction of the speech acquisition device **227**, the presence of the

influence of the environmental sound, or the deterioration of the communication) to the text data; and

[0290] A function of correcting a portion associated with a predetermined determination result (the malfunction of the speech acquisition device **227**, the presence of the influence of environmental sound, or the deterioration of the communication) in the text data, using text data before and after the portion associated with the predetermined determination result.

[0291] In FIG. **15**, a reference numeral **1520** denotes an example of the text data stored in the text data storage unit **450**. As illustrated in FIG. **15**, the text data stored in the text data storage unit **450** is stored in such a manner that it is possible to identify which one of the Web terminal **1420** and the Web terminal **1430** converted the audio data into the text data.

[0292] In FIG. **9**, a reference numeral **1530** indicates a state in which an area surrounded by a dotted line, among the text data indicated by the reference numeral **1520**, is added with marks by the text data correction unit **1510** to visualize the determination results.

[0293] Among these determination results, a reference numeral **1531** indicates the text data for a case where it is determined that the speech data can be converted into the text data, or that the quality of the converted text data after the conversion does not deteriorate, as a result of being notified of the "no influence of environmental sound", "normal operating state of microphone", and "normal communication".

[0294] On the other hand, a reference numeral **1532** indicates the text data after the addition of the mark for a case where it is determined that the speech data cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the influence of the environmental sound as a result of being notified of the "presence of influence of environmental sound". As indicated by the reference numeral **1532**, the text data correction unit **1510** adds a mark according to the cause to a corresponding position of the text data.

[0295] Similarly, a reference numeral **1533** indicates the text data after the addition of the mark for a case where it is determined that the speech data cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the malfunction of the speech acquisition device **227** as a result of being notified of the "abnormality of microphone". As indicated by the reference numeral **1533**, the text data correction unit **1510** adds a mark according to the cause to a corresponding position of the text data.

[0296] Similarly, a reference numeral **1534** indicates the text data after the addition of the mark for a case where it is determined that the speech data cannot be converted into the text data, or that the quality of the converted text data after the conversion deteriorates, due to the deterioration of the communication as a result of being notified of the "deterioration of communication". As indicated by the reference numeral **1534**, the text data correction unit **1510** adds a mark according to the cause to a corresponding position of the text data.

[0297] As described above, in the third embodiment, the server apparatus **110** adds a mark according to the cause of the determination result to the text data. Thus, according to the third embodiment, the client terminal (the Web terminal **1420** or **1430**) can visualize the determination result associated with the text data when displaying the text data. As a result, according to the third embodiment, even in a case where an appropriate speech recognition is not performed on a portion of the text data, the users **1450** and **1450** can understand the cause thereof.

[0298] The description given heretofore relate to a specific example of the function of visualizing the cause of the determination result among the correction functions of the text data correction unit **1510**, but a correction function using text data before and after a portion associated with a predetermined determination result can:

[0299] Correct the text data indicated by the reference numerals **1532** through **1534**,

[0300] Into the text data indicated by the reference numeral **1531**, for example.

This is because the determination result is displayed during the communication, and it is thus possible to shorten a time range in which the situation unsuitable for the speech recognition occurs (it is possible to keep the time range within a correctable time range using the text data before and after the portion associated with the predetermined determination result).

Conclusion

[0301] As is clear from the description given heretofore, the bidirectional communication system **1400**, which is an example of the system according to the third embodiment, includes the same functions as those of the first or second embodiment, and further includes:

[0302] A function of visualizing the cause of the determination result associated with the text data; and

[0303] A function of correcting a portion associated with the determination result using the text data before and after the portion associated with the determination result.

[0304] Hence, according to the third embodiment, the same effects as those of the first or second embodiment are obtainable, and further, even in the case where an appropriate speech recognition is not performed on a portion of the text data, the cause of the malfunction can be understood. Further, the data can be corrected to an appropriate text data.

Fourth Embodiment

[0305] In the first and second embodiments described above, a description of a communication process for a case where video data is transmitted to the supporter terminal **130** is omitted, but the video data may be stored in association with the text data, for example. In this case, a work content of the worker **150** may be identified and converted into character information by analyzing the video data, and the identified work content may be stored in association with the text data, for example.

[0306] FIG. **16** is a third diagram illustrating an example of the details of the functional configuration of the server apparatus. The difference from the first diagram illustrated in FIG. **4** is the provision of a work content identification unit **1610**.

[0307] The work content identification unit **1610** analyzes the video data notified from the communication unit **410**, and identifies the work content to be performed by the worker **150**, thereby converting the video data into the character information. The work content identification unit

1610 stores the identified work content in the text data storage unit **450** in association with the text data and the determination result.

[0308] Accordingly, by storing the work content in association with the text data and the determination result, when the supporter **160** browses the text data, for example, the supporter **160** can check the speech content of the worker **150** after recognizing the work content to be performed by the worker **150**.

Fifth Embodiment

[0309] In the first and second embodiments, the worker terminal **120** notifies the worker **150** of the determination result via the speech output device **228**, as described above. However, the method of notifying the worker **150** of the determination result is not limited to the above. For example, an LED or the like may be provided on the worker terminal **120** to emit light in a color according to the determination result. Alternatively, a vibrator or the like may be provided on the worker terminal **120**, and the worker terminal **120** may be vibrated at a frequency or amplitude according to the determination result.

[0310] The first and second embodiments are described above for a case where the server apparatus **110** transmits a display instruction according to a determination result for the worker terminal **120**, and a display instruction according to a determination result for the supporter terminal **130**, to the supporter terminal **130**. However, the instruction transmitted by the server apparatus **110** is not limited to the display instruction according to the determination result, and a transmission destination is not limited to the supporter terminal **130**. For example, the server apparatus **110** may be configured to transmit a notification instruction according to the determination result to the worker terminal **120**.

[0311] In each of the embodiments described above, the influence of the environmental sound and the abnormality of the microphone are described as the cause of the inability to convert the audio data into the text data or the cause of the deterioration in the quality of converted text data after the conversion. However, the cause of the inability to convert the audio data into the text data or the cause of the deterioration in the quality of the converted text data after the conversion is not limited to the influence of the environmental sound or the abnormality of the microphone, and for example, the cause may be an influence of audio data preceding the audio data being acquired on the audio data being acquired. This is because the audio data preceding the audio data being acquired may affect the audio data being acquired due to echo.

[0312] The presence or absence of the influence of the audio data preceding the audio data being acquired on the audio data being acquired can be determined by the same mechanism that determines the presence or absence of the influence of the environmental sound. For example, the presence or absence of the influence on the audio data being acquired can be determined by separating the audio data preceding the audio data being acquired from the audio data being acquired, and calculating a magnitude (a sound pressure level) of the audio data preceding the audio data being acquired.

[0313] A display method for displaying the presence or absence of the influence on the audio data being acquired may be the same as, or different from, the display method for displaying the presence or absence of the influence of the environmental sound.

[0314] Moreover, in each of the embodiments described above, the state of the environmental sound is detected by determining whether or not the sound pressure level of the environmental audio data notified from the sound pressure level calculation unit **330** is greater than or equal to the predetermined threshold value (that is, an absolute value). However, the method of detecting the state of the environmental sound is not limited thereto, and for example, the state of the environmental sound may be detected by determining whether or not a ratio between the magnitude (the sound pressure level) of the audio data and the magnitude (the sound pressure level) of the environmental audio data satisfies a predetermined condition (that is, a relative value). Specifically, in a case where the sound pressure level of the audio data is less than the sound pressure level of the environmental audio data, it may be determined that there is presence of the influence of the environmental sound may be determined, and in a case where the sound pressure level of the audio data is greater than or equal to the sound pressure level of the environmental audio data, it may be determined that there is no influence of the environmental sound.

[0315] Further, in each of the embodiments described above, the details of the method of acquiring the audio data by the audio data acquisition units **310** and **510** are not described. However, the audio data acquisition units **310** and **510** may be configured to perform various kinds of preprocesses when acquiring the audio data. The preprocess may include a process of removing a high-frequency component from the audio data using a lowpass filter, for example, and a process of removing hum noise by removing a low-frequency component from the audio data, for example. Alternatively, the preprocessing may include a process of canceling howling, echo, or the like by analyzing a waveform for each predetermined section, for example.

[0316] In addition, in each of the embodiments described above, the situation of the worker terminal **120** and the situation of the supporter terminal **130** are displayed using the icons illustrated in FIG. **8**. However, the method of displaying the status of the worker terminal **120** and the status of the supporter terminal **130** is not limited thereto, and the statuses may be displayed in other arbitrary display modes.

[0317] Moreover, in the third embodiment, the mark illustrated in FIG. **15** is used as the mark indicating the cause of the predetermined determination result added to the text data. However, the method of displaying the mark indicating the cause of the predetermined determination result added to the text data is not limited thereto, and the mark may be displayed in other arbitrary display modes.

[0318] Further, in each of the embodiments described above, the client terminal transmits the audio data to the server apparatus **110** in real time. However, the configuration of the client terminal is not limited thereto. For example, the client terminal may be configured to transmit the audio data in real time and successively store the audio data, and in a case where a request is made from the server apparatus **110**, to transmit the audio data in a range according to the request to the server apparatus **110**.

[0319] In addition, the first and second embodiments describe the case where the text data generation unit is implemented in the server apparatus **110**, and the case where

the text data generation unit is implemented in the client terminals (the worker terminal **120** and the supporter terminal **130**). However, a variation of the functional allocation between the server apparatus **110** and the client terminals (the worker terminal **120** and the supporter terminal **130**) is not limited thereto. For example, the determination unit **430** of the server apparatus **110** may be implemented in one or a plurality of client terminal (the worker terminal **120** and/or the supporter terminal **130**).

[0320] Moreover, the text data correction unit **1510** described in the third embodiment may be implemented in the bidirectional communication system **100** described in the first embodiment. In this case, when the text data is downloaded and displayed, the supporter terminal **130** can visualize the determination result associated with the text data. As a result, according to the fifth embodiment, even in the case where an appropriate speech recognition is not performed in a portion of the text data, the supporter **160** can understand the cause thereof.

[0321] Further, in the first embodiment, the case where the wearable terminal is used as the worker terminal **120** is described, however, a portable terminal may be used in place of the wearable terminal.

[0322] In addition, in each of the embodiments described above, the server apparatus **110** is described as executing the bidirectional communication program by itself. However, in a case where the server apparatus **110** is configured by one or a plurality of computers, for example, the bidirectional communication program may be installed in each of the one or plurality of computers and executed in a form of distributed computing.

[0323] Moreover, in each of the embodiments described above, the method of downloading and installing the client program via the network is described as an example of the method of installing the client program. Although a source of the program download is not particularly mentioned, in the case of installing the program by such a method, the source of the program download may be a server apparatus accessibly storing the client program, for example. The server apparatus may be an apparatus on a cloud, receiving an access from the client terminal (the supporter terminal **130**, the Web terminal **1420**, or the Web terminal **1430**) via a network, and downloading the client program on condition of payment. That is, the server apparatus may be an apparatus on the cloud, providing a client program providing service.

[0324] Although the embodiments are described above, it will be understood that various modifications may be made in form and detail without departing from the subject matter and scope of the claims.

[0325] This application is based upon and claims priority to Japanese Patent Application No. 2022-140702 filed on Sep. 5, 2022, the entire contents of which are incorporated herein by reference.

DESCRIPTION OF REFERENCE NUMERALS

[0326] **100**: bidirectional communication system
[0327] **110**: server apparatus
[0328] **111**: bidirectional communication unit
[0329] **112**: text data management unit
[0330] **120**: worker terminal
[0331] **130**: supporter terminal
[0332] **310**: audio data acquisition unit
[0333] **320**: speech separation unit
[0334] **330**: sound pressure level calculation unit
[0335] **340**: first detection unit
[0336] **350**: second detection unit
[0337] **360**: communication unit
[0338] **370**: determination unit
[0339] **380**: speech output unit
[0340] **410**: communication unit
[0341] **420**: speech data acquisition unit
[0342] **430**: determination unit
[0343] **440**: text data generation unit
[0344] **510**: audio data acquisition unit
[0345] **520**: speech separation unit
[0346] **530**: sound pressure level calculation unit
[0347] **540**: first detection unit
[0348] **550**: second detection unit
[0349] **560**: communication unit
[0350] **570**: speech output unit
[0351] **580**: display control unit
[0352] **590**: determination unit
[0353] **800**: display screen
[0354] **910**: text data generation unit
[0355] **920**: communication unit
[0356] **1010**: communication unit
[0357] **1110**: text data generation unit
[0358] **1120**: communication unit
[0359] **1400**: bidirectional communication system
[0360] **1420**: Web terminal
[0361] **1430**: Web terminal
[0362] **1510**: text data correction unit
[0363] **1610**: work content identification unit

1. A system comprising: a worker terminal used by a worker; a supporter terminal used by a supporter who remotely supports the worker; and a server apparatus configured to perform a bidirectional communication between the worker terminal and the supporter terminal, wherein a control unit included in the server apparatus converts speech data into text data,

the control unit included in the server apparatus is configured to:

determine, before a communication ends, whether or not the speech data being acquired during the communication by a control unit included in the worker terminal can be converted into the text data; and

determine that the speech data acquired during the communication cannot be converted into the text data due to a deterioration of a communication state of the bidirectional communication, in a case where the speech data is not transmitted from the worker terminal for a certain period of time and the bidirectional communication is determined as being disconnected, and

a control unit included in the supporter terminal outputs a determination result.

2-12. (canceled)

13. The system as claimed in claim **1**, wherein the control unit included in the supporter terminal displays the determination result using an icon indicating a cause the speech data cannot be converted into the text data, or a cause the quality of the converted text data after the conversion deteriorates.

14. The system as claimed in claim **1**, wherein the control unit included in the worker terminal is configured to:

determine, during the acquisition of the speech data, whether or not the speech data being acquired can be

converted into the text data, or whether or not the quality of the converted text data after the conversion the speech data being acquired deteriorates; and

notify the determination result by at least one of sound, light, and vibration.

15. The system as claimed in claim **1**, wherein the control unit included in the server apparatus stores the text data and the determination result in association with each other.

16. The system as claimed in claim **15**, wherein the control unit included in the supporter terminal reads the stored text data and displays the read text data in a display mode according to the determination result.

17. The system as claimed in claim **15**, wherein the control unit included in the server apparatus corrects a portion that cannot be converted into the text data or a portion where the quality deteriorates, using text data before and after the portion that cannot be converted into the text data or before and after the portion where the quality deteriorates.

18. The system as claimed in claim **15**, wherein the control unit included in the server apparatus identifies a work content of the worker from an image captured by the worker terminal, and stores the identified work content in association with the determination result.

19. The system as claimed in claim **15**, wherein the control unit included in the worker terminal stores the speech data acquired during the communication, and in a case where a request is made from the server apparatus, transmits the speech data in a range according to the request to the server apparatus.

20. (canceled)

21. A server apparatus comprising a control unit configured to perform a bidirectional communication between a worker terminal used by a worker and a supporter terminal used by a supporter who remotely supports the worker,

wherein the control unit is configured to:

convert speech data into text data;

determine, before a communication ends, whether or not the speech data being acquired during the communication by a control unit included in the worker terminal can be converted into the text data;

determine that the speech data acquired during the communication cannot be converted into the text data due to a deterioration of a communication state of the bidirectional communication, in a case where the speech data is not transmitted from the worker terminal for a certain period of time and the bidirectional communication is determined as being disconnected; and

transmit an instruction according to a determination result to the worker terminal or the supporter terminal.

22. A program causing a control unit of a server apparatus configured to perform a bidirectional communication between a worker terminal used by a worker and a supporter terminal used by a supporter who remotely supports the worker, to perform a process including the steps of:

converting speech data into text data;

determining, before a communication ends, whether or not the speech data being acquired during the communication by a control unit included in the worker terminal can be converted into the text data;

determining that the speech data acquired during the communication cannot be converted into the text data due to a deterioration of a communication state of the bidirectional communication, in a case where the speech data is not transmitted from the worker terminal for a certain period of time and the bidirectional communication is determined as being disconnected; and

transmitting an instruction according to a determination result to the worker terminal or the supporter terminal.

23. The system as claimed in claim **1**, wherein the control unit included in the server apparatus is configured to:

determine, before the communication ends, whether or not a quality of the converted text data after the conversion of the speech data acquired during the communication by the control unit included in the worker terminal deteriorates, and

determine that the quality of the text data after the conversion deteriorates due to an influence of an environmental sound around a speech acquisition device included in the worker terminal, in a case where a magnitude of environmental audio data acquired during the communication is greater than or equal to a predetermined threshold value, or in a case where a ratio between a magnitude of the speech data and the magnitude of the environmental audio data acquired during the communication satisfies a predetermined condition.

24. The system as claimed in claim **1**, wherein the control unit included in the server apparatus is configured to:

determine, before the communication ends, whether or not the speech data being acquired during the communication by the control unit included in the worker terminal can be converted into the text data, and

determine that the speech data cannot be converted into the text data due to a malfunction of a speech acquisition device included in the worker terminal, in a case where both a magnitude of the speech data being acquired and a magnitude of the environmental audio data acquired during the communication are less than a predetermined threshold value.

* * * * *