

(19) Japan Patent Office (JP)

(12) Gazette of Unexamined
Patent Applications (A)(11) Publication Number
2004-145025

(P2004-145025A)

(43) Publication Date May 20, 2004 (2004.5.20)

(51) Int.Cl. ⁷	FI		Theme Codes (Ref.)
G10L 17/00	G10L 3/00	545C	5D015
G10L 15/20	G10L 3/00	545D	
G10L 21/02	G10L 3/02	301D	

Examination Request Not Yet Received No. of Claims 5 OL (Total of 25 Pages)

(21) Application No. 2002-310156 (P2002-310156)
 (22) Filing Date October 24, 2002 (2002.10.24)

(71) Applicant 598098526
 Aruze Co., Ltd.
 3-1-25, Ariake, Koto-ku, Tokyo
 (74) Agent 100106002
 Attorney Masayuki SHOBAYASHI
 (72) Inventor
 Atsushi FUJIMOTO
 Aruze Co., Ltd.
 3-1-25, Ariake, Koto-ku, Tokyo
 F Terms (Reference) 5D015 AA03

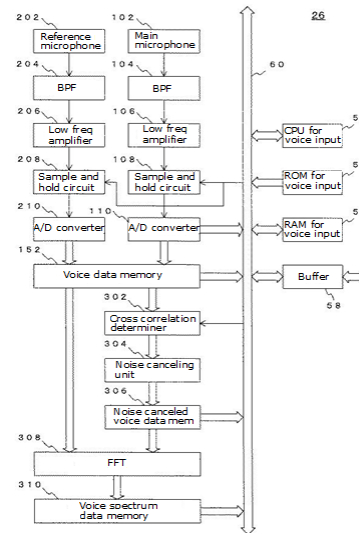
(54) [Title of the Invention] Voice Matching Device

(57) [Abstract]

[Problem] To provide a voice matching device with an increased likelihood of attaining a match.

[Solution] Provided is a voice matching device 10 comprising: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means.

[Selected Drawing] FIG. 4



[Claims]

[Claim 1]

A voice matching device that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means.

[Claim 2]

A voice matching device that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice recognizing means that performs voiceprint analysis and/or phonological analysis based on inputted voice signals, the device having a function of comparing voice signals of registered speakers with voice signals of the speaker to be matched based on the results of a voiceprint analysis and/or phonological analysis performed by the voice matching means; a voice comparing means that compares a plurality of voice signals inputted from each of the plurality of voice inputting means with each other, if the results of voiceprint analysis and/or phonological analysis based on voice signals inputted from one of the input devices do not yield a successful match; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; and a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means, wherein the voice recognizing means performs voiceprint analysis and/or phonological analysis based on one of the at least two approximate voice signals corrected by the voice correcting means, the device having a function to re-compare voice signals of registered speakers with the speaker to be matched based on the results of a voiceprint analysis and/or phonological analysis performed by the voice recognizing means.

[Claim 3]

A voice matching device that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a question information storing means for storing question information provided to the speaker to be matched; a question information providing means for providing the question information stored in the question information storing means to the speaker to be matched; a plurality of voice inputting means for inputting voice signals of the speaker to be matched as answer information to the question information provided by the question information providing means; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means.

[Claim 4]

A voice matching device that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a question information storing means for storing question information provided

to the speaker to be matched; an authentication information storing means for storing authentication information for the question information stored by the question information storing means; a question information providing means for providing the question information stored in the question information storing means to the speaker to be matched; a plurality of voice inputting means for inputting voice signals of the speaker to be matched as answer information to the question information provided by the question information providing means; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means, and authentication information stored by the authentication information storing means.

[Claim 5]

A voice matching device communicatively connected to the server that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means; a voice recognition results providing means for providing the results of a voiceprint and/or phonological analysis performed by the voice recognition means to the server; and a matching result receiving means by which the server receives the results of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of voiceprint analysis and/or phonological analysis provided by the means for providing speech recognition results.

[Detailed Description of the Invention]

[0001]

[Technical Field of the Invention]

The present invention relates to a voice matching device that performs matching by inputting voice data.

[0002]

[Prior Art]

A voice matching device usually performs matching based on the voice signals of registered speakers and voice signals of a speaker to be matched, and performs some authentication process based on the results of the matching process.

[0003]

Such a device digitizes inputted voice signals and performs predetermined processing on the digitized voice data to extract various voice elements such as phonemes and voiceprints, which can then be used to perform voice recognition.

[0004]

A voice matching device that performs identity matching by voice has been proposed (Patent Document 1) which enables voice matching with a high matching rate even when short voice signal data is used.

[0005]

[Patent Document 1] JP 2002-23792 A

[0006]

[Problem to Be Solved by the Invention]

However, such devices do not always successfully match voice signals. This is only possible if the inputted voice signals are of good quality, which requires various factors to be met in addition to voice data, such as capturing good quality voice signals and removing distortions from required voice signals. However, these devices do not yet sufficiently meet these factors, and voice signals are not always matched successfully.

[0007]

These devices can analyze voiceprints and identify speakers to be matched, but there is a risk that recorded voice signals of the speaker to be matched may be inputted by a person who is impersonating the speaker to be matched, and it is hoped that such a problem can be solved.

[0008]

It is an object of the present invention to solve this problem by providing a voice matching device with an increased likelihood of attaining a match.

[0009]

[Means for Solving the Problem]

In order to realize this object, the present invention is a voice matching device that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means.

[0010]

More specifically, the present invention provides the following.

[0011]

(1) A voice matching device that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological

analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means.

[0012]

Because the invention according to (1) "comprises: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[0013]

(2) A voice matching device that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice recognizing means that performs voiceprint analysis and/or phonological analysis based on inputted voice signals, the device having a function of comparing voice signals of registered speakers with voice signals of the speaker to be matched based on the results of a voiceprint analysis and/or phonological analysis performed by the voice matching means; a voice comparing means that compares a plurality of voice signals inputted from each of the plurality of voice inputting means with each other, if the results of voiceprint analysis and/or phonological analysis based on voice signals inputted from one of the input devices do not yield a successful match; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; and a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means, wherein the voice recognizing means performs voiceprint analysis and/or phonological analysis based on one of the at least two approximate voice signals corrected by the voice correcting means, the device having a function to re-compare voice signals of registered speakers with the speaker to be matched based on the results of a voiceprint analysis and/or phonological analysis performed by the voice recognizing means.

[0014]

Because the invention according to (2) "comprises: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice recognizing means that performs voiceprint analysis and/or phonological analysis based on inputted voice signals, the device having a function of comparing voice signals of registered speakers with voice signals of the speaker to be matched based on the results of a voiceprint analysis and/or phonological analysis performed by the voice matching means; a voice comparing means that compares a plurality of voice signals inputted from each of the plurality of voice inputting means with each other, if the results of voiceprint analysis and/or phonological analysis based on voice signals inputted from one of the input devices do not yield a successful match; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; and a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means, wherein the voice recognizing means performs voiceprint analysis and/or phonological analysis based on one of the at least two approximate voice signals corrected by the voice correcting means, the device having a function to re-compare voice

signals of registered speakers with the speaker to be matched based on the results of a voiceprint analysis and/or phonological analysis performed by the voice recognizing means," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[0015]

Also, in the event of a match failure, because this noise is removed by correction, high-quality voice signals are collected automatically as needed without any extra processing, and the cases of efficient processing performed by the voice matching device increases.

[0016]

(3) A voice matching device that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a question information storing means for storing question information provided to the speaker to be matched; a question information providing means for providing the question information stored in the question information storing means to the speaker to be matched; a plurality of voice inputting means for inputting voice signals of the speaker to be matched as answer information to the question information provided by the question information providing means; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means.

[0017]

Because the invention according to (3) "comprises: a question information storing means for storing question information provided to the speaker to be matched; a question information providing means for providing the question information stored in the question information storing means to the speaker to be matched; a plurality of voice inputting means for inputting voice signals of the speaker to be matched as answer information to the question information provided by the question information providing means; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[0018]

This voice recognizing means compares voiceprint information of registered speakers with voiceprint information of the speaker to be matched. Therefore, if more voice data can be inputted, a voice matching device can be provided with an even greater chance of a successful match.

[0019]

Also, a plurality of questions to be provided to the speaker to be matched can be stored in the question information storing means, and a question determining means may be provided for determining the piece of question information to be provided to the speaker to be matched from the plurality of pieces of question information stored in the question information storing means. In this case, various types of questions can be provided to the speaker to be matched, and various types of answer information can be obtained based on these types of questions. This enables more voice data to be inputted and provides a voice matching device with an even greater chance of a successful match.

[0020]

(4) A voice matching device that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a question information storing means for storing question information provided to the speaker to be matched; an authentication information storing means for storing authentication information for the question information stored by the question information storing means; a question information providing means for providing the question information stored in the question information storing means to the speaker to be matched; a plurality of voice inputting means for inputting voice signals of the speaker to be matched as answer information to the question information provided by the question information providing means; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means, and authentication information stored by the authentication information storing means.

[0021]

Because the invention according to (4) "comprises: a question information storing means for storing question information provided to the speaker to be matched; an authentication information storing means for storing authentication information for the question information stored by the question information storing means; a question information providing means for providing the question information stored in the question information storing means to the speaker to be matched; a plurality of voice inputting means for inputting voice signals of the speaker to be matched as answer information to the question information provided by the question information providing means; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means, and authentication information stored by the authentication information storing means," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[0022]

This voice recognizing means compares voiceprint information of registered speakers with voiceprint information of the speaker to be matched. Therefore, if more voice data can be inputted, a voice matching device can be provided with an even greater chance of a successful match.

[0023]

Also, a plurality of questions to be provided to the speaker to be matched can be stored in the question information storing means, and a question determining means may be provided for determining the piece of question information to be provided to the speaker to be matched from the plurality of pieces of question information stored in the question information storing means. In this case, various types of questions can be provided to the speaker to be matched, and various types of answer information can be obtained based on these types of questions. This enables more voice data to be inputted and provides a voice matching device with an even greater chance of a successful match.

[0024]

Also, the voice recognition means not only compares and matches registered and stored phonemes with the phonemes of the speaker to be matched, but also compares and matches registered and stored voiceprint information of registered speakers with voiceprint information of the speaker to be matched.

[0025]

Therefore, if a speaker to be matched different from a registered speaker inputs answer information during matching, the information will not be matched, thus preventing possible "spoofing" and increasing security.

[0026]

(5) A voice matching device communicatively connected to the server that compares voice signals of registered speakers with those of a speaker to be matched, comprising: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means; a voice recognition results providing means for providing the results of a voiceprint and/or phonological analysis performed by the voice recognition means to the server; and a matching result receiving means by which the server receives the results of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of voiceprint analysis and/or phonological analysis provided by the means for providing speech recognition results.

[0027]

Because the invention according to (5) "comprises: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means; a voice recognition results providing means for providing the results of a voiceprint and/or phonological analysis performed by the voice recognition means to the server; and a matching result receiving means by which the server receives the results of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of voiceprint analysis and/or phonological analysis provided by the means for providing speech recognition results," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This

makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[0028]

Also, because a plurality voice matching devices can be connected to such a server, a speaker can be registered as a registered speaker with the server and matching can be performed against voice matching devices connected to the server, making the process simple and convenient.

[0029]

[Embodiment of the Invention]

A preferred embodiment of the present invention will now be described based on the drawings.

[0030]

[Configuration of the Voice Matching Device]

FIG. 1 is a block diagram of the voice matching device.

[0031]

As shown in FIG. 1, the voice matching device 10 consists of a main CPU 21, a main memory 22, a display unit 23 consisting of a liquid crystal display panel or CRT, etc., an input unit 24 consisting of a keyboard, jog dial, etc., a communication interface (communication I/F) 25, a voice input device 26, and a storage unit 27 consisting of a hard disk and a hard disk drive, etc. connected to each other via a databus BUS.

[0032]

The main CPU 21 is configured to execute various operations according to an operation program stored in the main memory 22, and controls each circuit unit according to these operations. Details of the processing performed by the main CPU 21 are displayed on the display unit 23 as necessary.

[0033]

In addition, as described below, a server 81, etc. can be connected to the communication I/F 25, and the matching process may be performed via the server 81.

[0034]

[Configuration of the Storage Unit in the Voice Matching Device]

The storage unit 27 also stores a database, as shown in FIG. 2.

[0035]

A database of registered speakers is stored in this storage unit 27 as shown in FIG. 2.

[0036]

This registered speaker database is where information on registered speakers who use the voice matching device 10 is associated and stored, and these items are associated with information such as an "ID number," "question ID," "answer information," "voiceprint information ID," etc.

[0037]

This "question ID" indicates the question number, and as shown in FIG. 3, the content of the question is determined by referring to the question table that associates the "question ID" with "question information" stored in the storage unit 27.

[0038]

The "answer information" corresponds to the "question ID," and is registered by registered speakers when they register to use the system. This information is registered based on voice signals inputted from the voice input device 26 described below or is registered based on operating the input unit 24.

[0039]

Note that the "question information" and the "answer information" associated with a "question ID" may be associated in a way that makes sense, or may be associated in a way that makes no sense. For example, if the "question information" is "Where are you from?," then the answer information could be "Gifu Prefecture" (which has meaning) or "Wednesday" (which has no meaning).

[0040]

"Voiceprint information ID" is a number that indicates voiceprint information, and the contents of the voiceprint information are determined by referring to the voiceprint information table that associates the "voiceprint information ID" with "voiceprint information" stored in the storage unit 27. This voiceprint information table is stored in the storage unit 27 and registered by a registered speaker when they registered to use the system. It is registered based on voice signals inputted from the voice input device 26 (described later) or based on operating the input unit 24.

[0041]

[Configuration of the Voice Input device]

FIG. 4 is a block diagram showing the voice input device of the present invention.

[0042]

The voice input device 26 is equipped with a main microphone 102 and a reference microphone 202, as shown in FIG. 4. Note that while FIG. 4 shows a single reference microphone 202, in reality, multiple reference microphones are provided, including this reference microphone 202.

[0043]

This main microphone 102 receives main voice signals ("main voice signals" below).

[0044]

The reference microphone 202 differs from the main microphone 102, which receives the main voice signals, in that it receives voice signals ("reference voice signals" below) that are used to correct the main voice signals supplied by the main microphone 102.

[0045]

The main microphone 102 is connected to a bandpass filter (BPF) 104. This BPF 104 is a filter that only allows signals within a specified frequency band to pass through, that is, the frequency band of the sounds emitted by the speaker, regardless of whether they are male or female, and can attenuate signals in frequency bands different from that frequency band in advance. Therefore, it receives the main voice signals supplied from the main microphone 102, filters the main voice signals, and then outputs them.

[0046]

A bandpass filter (BPF) 204 is also connected to the reference microphone 202.

[0047]

This BPF 104 is a filter that only allows signals within a specified frequency band to pass through, that is, the frequency band of the sounds emitted by the speaker, regardless of whether they are male or female, and can attenuate signals in frequency bands different from that frequency band in advance. Therefore, it receives reference voice signals supplied from the reference microphone 202, filters the reference voice signals, and then outputs them.

[0048]

A low-frequency amplifier 106 is connected to the BPF 104. This low-frequency amplifier 106 amplifies the specified low-frequency components of the main voice signals that have passed through the BPF 104. This process allows voice to be converted into signals that can be accurately recognized by voice recognition processing.

[0049]

A low-frequency amplifier 206 is also connected to the BPF 204. This low-frequency amplifier 206 amplifies the specified low-frequency components of the reference voice signals that have passed through the BPF 204. This process allows voice to be converted into signals that can be accurately recognized by voice recognition processing.

[0050]

This low-frequency amplifier 106 is connected to a sample and hold circuit 108. This sample and hold circuit 108 is connected to an input/output bus 60, and it is used to accurately determine the value of analog signals to be converted in response to control signals from the input/output bus 60.

[0051]

The input/output bus 60 is configured so that data signals or address signals are inputted and outputted to the CPU 52 for voice input.

[0052]

The input/output bus 60 is also connected to ROM 54 and RAM 56 for voice input. The ROM 54 for voice input stores the program used to perform various operations and the initial data used when executing the program.

[0053]

The RAM 56 for voice input stores the values of the flags and variables used during the execution of the program.

[0054]

The sample and hold circuit 108 is connected to the input/output bus 60, and the CPU 52 for voice input sends control signals to the sample and hold circuit 108 at predetermined intervals.

[0055]

A sample and hold circuit 208 is also connected to the low-frequency amplifier 206. In addition, the sample and hold circuit 208 is connected to the input/output bus 60, and is used to accurately determine the values of voice signals to be converted in response to control signals from the input/output bus 60.

[0056]

In addition, the sample and hold circuit 208 is connected to the input/output bus 60, and the CPU 52 for voice input sends control signals to the sample and hold circuit 108 at predetermined intervals.

[0057]

By using sample and hold circuits 108, 208 in this way, the main voice signals and reference voice signals are outputted on a specified timing even at the rise time of the voice signals, and synchronization of the main voice signals and reference voice signal can be realized.

[0058]

Sample and hold circuit 108 is connected to A/D converter 110. A/D converter 110 converts analog signals to digital signals, receiving the main voice signals from the sample and hold circuit 108, converting them to digital form, and outputting the main voice data.

[0059]

Sample and hold circuit 208 is connected to A/D converter 210. A/D converter 210 converts analog signals to digital signals, receiving the reference voice signals from the sample and hold circuit 208, converting them to digital form, and outputting the reference voice data.

[0060]

These A/D converters 110 and 210 are connected to the voice data memory 152. This voice data memory 152 is used to record the main voice data and reference voice data after various noise removal processes have been applied, such as filtering and amplification. In addition, the voice data memory 152 is also connected to the input/output bus 60, so that voice data may be read by the CPU 52 for voice input. Note that this voice data memory 152 records both the main voice data and the reference voice data.

[0061]

By performing this processing, voice signals inputted to the main microphone 102 and the reference microphone 202 can be handled as voice data.

[0062]

This voice data memory 152 is also connected to a cross correlation determiner 302. The cross correlation determiner 302 reads the main voice data supplied from the main microphone 102 and the reference data supplied from the multiple reference microphones including reference microphone 202 from the voice data memory 152, and cross compares the main voice data with the reference data.

[0063]

In addition, the cross correlation determiner 302 extracts that which is most similar to the main voice data from the reference data based on the results of the comparison, and outputs the main voice data and the extracted reference voice data ("extracted reference voice data" below).

[0064]

This cross correlation determiner 302 is connected to a noise canceling unit 304. This noise canceling unit 304 receives the main voice data supplied by the cross correlation determiner 302 and the extracted reference voice data, and by comparing them, corrects the main voice data and removes noise. Note that this noise canceling unit 304 corrects the differences between the main voice data and the extracted reference voice data.

[0065]

This noise canceling unit 304 is connected to a noise cancelled voice data memory 306. This noise-canceling voice data memory 306 records the main voice data after noise-canceling processing has been

applied to the main voice signals by the noise-canceling unit 304. In addition, this noise-canceling voice data memory 306 is connected to the input/output bus 60, so noise-canceling voice data may be read by the CPU 52 for voice input.

[0066]

This noise-canceling voice data memory 306 is connected to a fast Fourier transform processing unit (FFT) 308. The FFT 308 reads the main voice data recorded in the noise-canceling voice data memory 306, performs a fast Fourier transform, and outputs the converted data as voice spectrum data. In this way, voice spectrum data is extracted.

[0067]

The FFT 308 is also directly connected to the voice data memory 152, so it is also possible to perform a fast Fourier transform without the cross correlation and noise canceling processing described above.

[0068]

This FFT 308 is connected to a voice spectrum data memory 310. This voice spectrum data memory 310 records the voice spectrum data obtained by performing a fast Fourier transform on noise-canceled main voice data. In this way, voice signals captured by the main microphone 102 can be converted into a frequency/intensity relationship. In addition, the voice spectrum data memory 310 is connected to the input/output bus 60, and the voice data that has been subjected to a fast Fourier transform may be read by the CPU 52 for voice input.

[0069]

In addition, the CPU 52 for voice input, which receives voice data, noise-canceling voice data, and noise-canceling voice spectrum data, supplies it to the main CPU 21 via a buffer 58. The main CPU 21 receiving this data then performs voiceprint analysis and phonological analysis using the database and table shown in FIG. 2 and FIG. 3, and as a result performs the matching process.

[0070]

As mentioned above, the configuration of the buffer 58 allows large amounts of data, such as voice data and extracted voice data, to be supplied accurately to the main CPU 21.

[0071]

In addition, the voice input device 26 described above may be configured using a digital signal processor (DSP).

[0072]

[Explanation of the Operations Performed by the Voice Matching Device]

In the following explanation, the voice matching device 10 is running, and the variables used by the main CPU 21 and voice input CPU 52 have been initialized to the specified values and are operating normally.

[0073]

The subroutine executed by the main CPU 21 described above is shown in FIG. 5. This subroutine is called and executed at predetermined intervals.

[0074]

First, as shown in FIG. 5, it is determined whether or not a matching start signal has been received (step S11). In this process, the main CPU 21 determines that it has received a matching start signal if

an ID number is inputted based on the inputted signal from the input unit 24. In this process, if the main CPU 21 determines that it has received a matching start signal, the routine moves on to step S12, and if it does not determine that it has received a matching start signal, it ends the subroutine.

[0075]

Next, if it is determined that a matching start signal has been received based on the determination in step S11, the question information selection process is executed (step S12). In this process, the main CPU 21 randomly records a question ID in the main memory 22. The main CPU 21 then determines the question information based on the question ID recorded in the main memory 22 and the question table. If this process is complete, the routine moves on to step S13.

[0076]

Next, the question information providing process is executed (step S13). In this process, the main CPU 21 provides question information to the speaker based on the question ID recorded in the main memory 22 and the question table from the processing in step S12. In other words, the main CPU 21 displays an image showing the question information on the display unit 23, and makes the sound generator (not shown in the figure) emit a voice stating the question information. If this process is complete, the routine moves on to step S21.

[0077]

Next, voice input processing is executed using multiple microphones (step S21). In this process, the main microphone 102 and the reference microphone 202 each record voice, and supply the main voice signals and the reference voice signal, which indicate the collected voice data, to BPFs 104 and 204, respectively. If this process is complete, the routine moves on to step S22.

[0078]

Next, the voice signals are filtered (step S22). In this process, BPF 104 and 204 pass through only main voice signals and reference voice signals received from the main microphone 102 and reference microphone 202, respectively, in the specified frequency band. Signals in frequency bands different from the frequency band are attenuated in advance, and the filtered main voice signals and reference voice signals are supplied to low-frequency amplifiers 106 and 206. If this process is complete, the routine moves on to step S23.

[0079]

Next, low-frequency amplification processing is performed (step S23). In this process, the low-frequency amplifiers 106 and 206 amplify the specified low-frequency components of the main voice signals and reference voice signal supplied from BPFs 104 and 204. This process allows for the conversion of voice into signals that can be accurately recognized by voice recognition processing. The low-frequency amplifiers 106 and 206 supply the amplified main and reference voice signals to the sample and hold circuits 108 and 208. If this process is complete, the routine moves on to step S24.

[0080]

Next, the sample and hold processing is executed (step S24). In this process, the sample and hold circuits 108 and 208 supply the main voice signals and reference voice signals supplied by the low-frequency amplifiers 106 and 206 to the A/D converters 110 and 210 on a timing based on control signals supplied by the CPU 52 for voice input, thereby synchronizing the main voice signals and reference voice signals. If this process is complete, the routine moves on to step S25.

[0081]

Next, the A/D conversion processing is executed (step S25). In this process, A/D converters 110 and 210 receive the main voice signals and reference voice signals supplied by the sample and hold circuits

108 and 208, and convert them to digital signals. If this process is complete, the routine moves on to step S26.

[0082]

Next, the voice data recording process is executed (step S26). In this process, A/D converters 110 and 210 record the digitally converted main voice data and reference voice data in the voice data memory 152. When this process is complete, the routine moves to step S27.

[0083]

Next, as shown in FIG. 6, the cross correlation determination processing is executed (step S27). In this process, the cross correlation determiner 302 reads the main voice data supplied by the main microphone 102 and the reference data supplied by the multiple reference microphones, including reference microphone 202, from the voice data memory 152. The cross correlation determiner 302 compares the main voice data with the reference data, and as a result of this comparison, extracts the reference data that is most similar to the main voice data from the reference data. The cross correlation determiner 302 then supplies the main voice data and the extracted reference voice data ("extracted reference voice data" below) to the noise canceling unit 304. When this processing is complete, the routine moves on to step S28.

[0084]

Next, the noise cancellation processing is executed (step S28). In this process, the noise canceling unit 304 receives the main voice data and the extracted reference voice data supplied by the cross correlation determiner 302 and compares them. The noise canceling unit 304 then corrects the main voice data so that the difference between the main voice data and the extracted reference voice data is reduced. When this processing is complete, the routine moves on to step S29.

[0085]

Next, the voice data recording processing is executed (step S29). In this process, the noise canceling unit 304 records the main voice data that has been noise-canceled by the process in step S28 in the noise-canceling voice data memory 306. When this processing is complete, the routine moves on to step S30.

[0086]

Next, the voice data is processed using a fast Fourier transform (step S30). In this process, the FFT 308 reads the main voice data recorded in the noise-canceling voice data memory 306 and performs a fast Fourier transform on the main voice data. When this process is complete, the routine moves on to step S31.

[0087]

Next, the voice spectrum data recording processing is executed (step S31). In this process, the FFT 308 records the voice spectrum data converted by the process in step S30 in the voice spectrum data memory 310. When this process is finished, the routine moves on to step S32.

[0088]

Next, it is determined whether a predetermined amount of time has elapsed (step S32). In this process, the main CPU 21 uses the timer built into the main CPU 21 to determine whether the predetermined amount of time has elapsed since the question information was provided. In this process, if the main CPU 21 determines that the predetermined amount of time has elapsed, it moves on to step S41, and if it does not determine that the predetermined amount of time has elapsed, it moves on to step S21.

[0089]

Next, if it is determined that the predetermined amount of time has elapsed based on the determination in Step S32, the voice spectrum data extraction process is executed (step S41). In this process, the main CPU 21 instructs the CPU 52 for voice output to read out the voice spectrum data. The CPU 52 for voice input, which has received this instruction, reads the voice spectrum data recorded in the voice spectrum data memory 310 and supplies it to the main CPU 21. The main CPU 21 then extracts voiceprint information from the voice spectrum data. When this processing is complete, the routine moves on to step S42.

[0090]

Next, the answer information generating process is executed (step S42). In this process, the main CPU 21 performs a phonological analysis on the voice spectrum data received in step S41 to generate answer information. When this process is complete, the routine moves on to step S43.

[0091]

Next, the database reference process is executed (step S43). In this process, the main CPU 21 references the registered voice database recorded in the storage unit 27. In this process, the main CPU 21 reads predetermined data in the registered speaker database based on the ID number entered in step S11 and the question number determined in step S12. If this process is complete, the routine moves on to step S44.

[0092]

Next, the voice matching process is executed (step S44). In this process, the main CPU 21 compares the voiceprint information extracted in step S41 with the reference results from step S43, and determines whether they are similar or not. The main CPU 21 also compares the phoneme information generated by the processing in Step S42 with the reference results from the processing in Step S43, and determines whether they are similar or not. In this way, the main CPU 21 establishes a matching result by successfully matching voiceprint information and phoneme information.

[0093]

In the present embodiment, the voiceprint information, phoneme information, and matching of these are successfully carried out, and this is an ideal example in which the matching result is successful. However, the present invention is not limited to this, and it is also acceptable for matching results to be established by the successful matching of either voiceprint information or phoneme information, and it is preferable for the matching result to be established by the successful matching of voiceprint information. If this process is complete, the routine moves on to step S51.

[0094]

Next, as shown in FIG. 7, it is determined whether or not a match has been established (step S51). In this process, the main CPU 21 makes a determination based on the results of the matching process in step S44. In this process, if the main CPU 21 determines that a match has been established, the process moves to step S52, and if it determines that a match has not been established, the process moves to step S53.

[0095]

Next, when it has been determined that a match has been established according to the determination in step S51, successful match display processing is executed (step S52). In this process, the main CPU 21 displays an image on the display unit 23 to indicate that a match has been established. When this process is completed, the subroutine ends.

[0096]

If it has been determined by the determination in step S51 that a match has not been established, failed match display processing is executed (step S53). In this process, the main CPU 21 displays an image

on the display unit 23 to indicate that a match was not established. When this process is completed, the subroutine ends.

[0097]

Because an embodiment "comprises: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[0098]

Because an embodiment also "comprises: a question information storing means for storing question information provided to the speaker to be matched; a question information providing means for providing the question information stored in the question information storing means to the speaker to be matched; a plurality of voice inputting means for inputting voice signals of the speaker to be matched as answer information to the question information provided by the question information providing means; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[0099]

This voice recognizing means compares voiceprint information of registered speakers with voiceprint information of the speaker to be matched. Therefore, if more voice data can be inputted, a voice matching device can be provided with an even greater chance of a successful match.

[0100]

Also, a plurality of questions to be provided to the speaker to be matched can be stored in the question information storing means, and a question determining means may be provided for determining the piece of question information to be provided to the speaker to be matched from the plurality of pieces of question information stored in the question information storing means. In this case, various types of questions can be provided to the speaker to be matched, and various types of answer information can be obtained based on these types of questions. This enables more voice data to be inputted and provides a voice matching device with an even greater chance of a successful match.

[0101]

Because an embodiment also "comprises: a question information storing means for storing question information provided to the speaker to be matched; an authentication information storing means for storing authentication information for the question information stored by the question information storing means; a question information providing means for providing the question information stored in the question information storing means to the speaker to be matched; a plurality of voice inputting means for inputting voice signals of the speaker to be matched as answer information to the question information provided by the question information providing means; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means, and authentication information stored by the authentication information storing means," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[0102]

This voice recognizing means compares voiceprint information of registered speakers with voiceprint information of the speaker to be matched. Therefore, if more voice data can be inputted, a voice matching device can be provided with an even greater chance of a successful match.

[0103]

Also, a plurality of questions to be provided to the speaker to be matched can be stored in the question information storing means, and a question determining means may be provided for determining the piece of question information to be provided to the speaker to be matched from the plurality of pieces of question information stored in the question information storing means. In this case, various types of questions can be provided to the speaker to be matched, and various types of answer information can be obtained based on these types of questions. This enables more voice data to be inputted and provides a voice matching device with an even greater chance of a successful match.

[0104]

Also, the voice recognition means not only compares and matches registered and stored phonemes with the phonemes of the speaker to be matched, but also compares and matches registered and stored voiceprint information of registered speakers with voiceprint information of the speaker to be matched.

[0105]

Therefore, if a speaker to be matched different from a registered speaker inputs answer information during matching, the information will not be matched, thus preventing possible "spoofing" and increasing security.

[0106]

In the present embodiment, noise cancellation processing is always performed. However, the present invention is not limited to this, and noise cancellation processing may be performed only when the matching result ends in a match failure.

[0107]

The operations performed in this case are explained using FIG. 8 to FIG. 10. In FIG. 8 to FIG. 10, the same numbers are used for steps that perform the same processing as those shown in FIG. 5 to FIG. 7.

[0108]

First, as shown in FIG. 8, it is determined whether or not a matching start signal has been received (step S11). In this process, the main CPU 21 determines that a matching start signal has been received if an ID number has been inputted based on inputted signals from the input unit 24. In this process, if the main CPU 21 determines that it has received a matching start signal, it moves on to step S16, and if it does not determine that it has received a matching start signal, it ends the subroutine.

[0109]

Next, if it is determined in Step S11 that a matching start signal has been received, the question information selection process is executed (step S16). In this process, the main CPU 21 reads the match failure count data located in the main memory 22 and records "00" as the data. This is because the match failure count is used to store the number of times a match was unsuccessful. When this process is complete, the process is moved to step S12.

[0110]

Next, the question information selection process is executed (step S12). In this process, the main CPU 21 randomly records the question ID in the main memory 22. The main CPU 21 then determines the question information based on the question ID recorded in the main memory 22 and the question table. When this process is complete, the routine moves on to step S13.

[0111]

Next, the question information provision process is executed (step S13). In this process, the main CPU 21 provides question information to the speaker based on the question ID and the question table recorded in the main memory 22 using the processing in step S12. In other words, the main CPU 21 displays an image showing question information on the display unit 23, and makes the sound generator (not shown in the figure) generate a voice stating the question information. When this processing is finished, the routine moves on to step S19.

[0112]

Next, it is determined whether or not the match was unsuccessful (step S19). In this process, the main CPU 21 reads the number of failed matches located in the main memory 22 and determines whether or not the data is "00". If the main CPU 21 determines that the match failed, it moves on to step S21, and if it does not determine that the match failed, it moves on to step S20.

[0113]

If the determination in step S19 indicates that the match was unsuccessful, voice input processing is executed using multiple microphones (step S21). In this processing, the main microphone 102 and reference microphone 202 each collect voice signals, and supply the main voice signals and reference voice signals, which represent the collected voice signals, to BPFs 104 and 204. If this processing is completed, the routine moves on to Step S22.

[0114]

If it is not determined that the match failed based on the determination in step S19, voice input processing is executed using the main microphone (step S20). In this processing, the main microphone 102 collects voice, and supplies the main voice signals indicating the collected voice signals to BPF 104. When this processing is completed, the routine moves on to Step S22.

[0115]

Next, the video signals are filtered (step S22). In this process, BPF 104 and 204 pass through only main voice signals and reference voice signals received from the main microphone 102 and reference microphone 202, respectively, in the specified frequency band. Signals in frequency bands different from the frequency band are attenuated in advance, and the filtered main voice signals and reference voice signals are supplied to low-frequency amplifiers 106 and 206. If this process is complete, the routine moves on to step S23.

[0116]

Next, low-frequency amplification processing is performed (step S23). In this process, the low-frequency amplifiers 106 and 206 amplify the specified low-frequency components of the main voice signals and reference voice signal supplied from BPFs 104 and 204. This process allows for the conversion of voice into signals that can be accurately recognized by voice recognition processing. The low-frequency amplifiers 106 and 206 supply the amplified main and reference voice signals to the sample and hold circuits 108 and 208. If this process is complete, the routine moves on to step S24.

[0117]

Next, the sample and hold processing is executed (step S24). In this process, the sample and hold circuits 108 and 208 supply the main voice signals and reference voice signals supplied by the low-frequency amplifiers 106 and 206 to the A/D converters 110 and 210 on a timing based on control signals supplied by the CPU 52 for voice input, thereby synchronizing the main voice signals and reference voice signals. If this process is complete, the routine moves on to step S25.

[0118]

Next, the A/D conversion processing is executed (step S25). In this process, A/D converters 110 and 210 receive the main voice signals and reference voice signals supplied by the sample and hold circuits 108 and 208, and convert them to digital signals. If this process is complete, the routine moves on to step S26.

[0119]

Next, the voice data recording process is executed (step S26). In this process, A/D converters 110 and 210 record the digitally converted main voice data and reference voice data in the voice data memory. When this process is complete, the routine moves to step S36.

[0120]

Next, as shown in FIG. 9, it is determined whether or not the match was unsuccessful (step S36). In this process, the main CPU 21 reads the number of failed matches located in the main memory 22 and determines whether or not the data is "00". If the main CPU 21 determines that the match failed, it moves on to step S27, and if it does not determine that the match failed, it moves on to step S30.

[0121]

If the match is determined to be unsuccessful based on the determination in Step S36, the cross correlation determination process is executed (step S27). In this process, the cross correlation determiner 302 reads the main voice data supplied by the main microphone 102 and the reference data supplied by the multiple reference microphones, including reference microphone 202, from the voice data memory 152. The cross correlation determiner 302 compares the main voice data with the reference data, and as a result of this comparison, extracts the reference data that is most similar to the main voice data from the reference data. The cross correlation determiner 302 then supplies the main voice data and the extracted reference voice data ("extracted reference voice data" below) to the noise canceling unit 304. When this processing is complete, the routine moves on to step S28.

[0122]

Next, the noise cancellation processing is executed (step S28). In this process, the noise canceling unit 304 receives the main voice data and the extracted reference voice data supplied by the cross correlation determiner 302 and compares them. The noise canceling unit 304 then corrects the main voice data so that the difference between the main voice data and the extracted reference voice data is reduced. When this processing is complete, the routine moves on to step S29.

[0123]

Next, the voice data recording processing is executed (step S29). In this process, the noise canceling unit 304 records the main voice data that has been noise-canceled by the process in step S28 in the noise-canceling voice data memory 306. When this processing is complete, the routine moves on to step S30.

[0124]

Next, the voice data is processed using a fast Fourier transform (step S30). In this process, under initial conditions in which there has not been a matching failure, the FFT 308 reads the main voice data recorded in the voice data memory 152 and performs a fast Fourier transform on the main voice data. However, under initial conditions in which there has been a matching failure, the FFT 308 reads the main voice data recorded in the noise-canceling voice data memory 306 and performs a fast Fourier transform on the main voice data. When this process is complete, the routine moves on to step S31.

[0125]

Next, the voice spectrum data recording processing is executed (step S31). In this process, the FFT 308 records the voice spectrum data converted by the process in step S30 in the voice spectrum data memory 310. When this process is finished, the routine moves on to step S32.

[0126]

Next, it is determined whether a predetermined amount of time has elapsed (step S32). In this process, the main CPU 21 uses the timer built into the main CPU 21 to determine whether the predetermined amount of time has elapsed since the question information was provided. In this process, if the main CPU 21 determines that the predetermined amount of time has elapsed, it moves on to step S41, and if it does not determine that the predetermined amount of time has elapsed, it moves on to step S21.

[0127]

Next, if it is determined that the predetermined amount of time has elapsed based on the determination in Step S32, the voice spectrum data extraction process is executed (step S41). In this process, the main CPU 21 instructs the CPU 52 for voice output to read out the voice spectrum data. The CPU 52 for voice input, which has received this instruction, reads the voice spectrum data recorded in the voice spectrum data memory 310 and supplies it to the main CPU 21. The main CPU 21 then extracts voiceprint information from the voice spectrum data. When this processing is complete, the routine moves on to step S42.

[0128]

Next, the answer information generating process is executed (step S42). In this process, the main CPU 21 performs a phonological analysis on the voice spectrum data received in step S41 to generate answer information. When this process is complete, the routine moves on to step S43.

[0129]

Next, the database reference process is executed (step S43). In this process, the main CPU 21 references the registered voice database recorded in the storage unit 27. In this process, the main CPU 21 reads predetermined data in the registered speaker database based on the ID number entered in step S11 and the question number determined in step S12. If this process is complete, the routine moves on to step S44.

[0130]

Next, the voice matching process is executed (step S44). In this process, the main CPU 21 compares the voiceprint information extracted in step S41 with the reference results from step S43, and determines whether they are similar or not. The main CPU 21 also compares the phoneme information generated by the processing in Step S42 with the reference results from the processing in Step S43, and determines whether they are similar or not. In this way, the main CPU 21 establishes a matching result by successfully matching voiceprint information, phoneme information, or both.

[0131]

The present embodiment is a good example of how matching results are established when voiceprint information, phoneme information, and both are successfully matched. However, the present invention is not limited to this, and matching results can be established by the successful matching of either voiceprint information or phoneme information. When this process is completed, the process moves to step S51.

[0132]

Next, as shown in FIG. 10, it is determined whether or not the matching has been established (step S51). In this process, the main CPU 21 makes a determination based on the result of the matching performed in step S44. In this process, if the main CPU 21 determines that a match has been established, the process moves to step S61, and if it determines that a match has not been established, the process moves to step S62.

[0133]

Next, when it has been determined that a match has been established in step S51, successful match recording processing is executed (step S62). The main CPU 21 reads the match failure count data located in the main memory 22 and records "00" as the data. When this process is completed, the process moves to step S52.

[0134]

Next, successful match display processing is executed (step S52). In this process, the main CPU 21 displays an image on the display unit 23 indicating that a match has been established. When this process is completed, the subroutine ends.

[0135]

If it is not determined in step S51 that a match has been established, the failed match recording processing is executed (step S62). In this process, the main CPU 21 reads the match failure count data located in the main memory 22, increments the data, and records the incremented data as match failure count data. When this process is completed, the process moves to step S63.

[0136]

Next, it is determined whether the number of match failures is less than or equal to a predetermined number (step S63). In this process, the main CPU 21 reads the match failure count data located in the main memory 22 and determines whether the data is less than or equal to a predetermined number. If the main CPU 21 determines that the number of match failures is less than or equal to the predetermined number, the process moves to step S13, and if it determines that the number of match failures is not less than the predetermined number, the process moves to step S53.

[0137]

Next, if the number of match failures has not been determined to be less than a predetermined number in step S63, the failed match display processing is executed (step S53). In this process, the main CPU

21 displays an image on the display unit 23 stating that a match has not been established. When this process is completed, the subroutine ends.

[0138]

Therefore, because an embodiment "comprises: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice recognizing means that performs voiceprint analysis and/or phonological analysis based on inputted voice signals, the device having a function of comparing voice signals of registered speakers with voice signals of the speaker to be matched based on the results of a voiceprint analysis and/or phonological analysis performed by the voice matching means; a voice comparing means that compares a plurality of voice signals inputted from each of the plurality of voice inputting means with each other, if the results of voiceprint analysis and/or phonological analysis based on voice signals inputted from one of the input devices do not yield a successful match; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; and a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means, wherein the voice recognizing means performs voiceprint analysis and/or phonological analysis based on one of the at least two approximate voice signals corrected by the voice correcting means, the device having a function to re-compare voice signals of registered speakers with the speaker to be matched based on the results of a voiceprint analysis and/or phonological analysis performed by the voice recognizing means," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[0139]

Also, in the event of a match failure, because this noise is removed by correction, high-quality voice signals are collected automatically as needed without any extra processing, and the cases of efficient processing performed by the voice matching device increases.

[0140]

The configuration of the present embodiment only includes a voice matching device 10, but the present invention is not limited to this, and a plurality of voice matching devices may be connected via a server.

[0141]

For example, as shown in FIG. 11, a plurality of voice matching devices 10 may be connected to a server 81, and the server 81 may perform matching based on information on voice signals supplied by a voice matching device 10 and send the results of the matching process to the voice matching device 10. In this case, the communication I/F 25 exchanges information with the server 81, and the registered speaker database for registered speakers is stored in the memory unit (not shown) of the server 81, allowing the server 81 to perform the matching process.

[0142]

Because when configured in this way, an embodiment "comprises: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means; a voice recognition results providing means for providing the results of a voiceprint and/or phonological analysis performed by the voice recognition means to the server; and a matching result receiving means by which the server receives the results of matching voice signals from registered

speakers with voice signals of the speaker to be matched based on the results of voiceprint analysis and/or phonological analysis provided by the means for providing speech recognition results," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[0143]

Also, because a plurality voice matching devices can be connected to such a server, a speaker can be registered as a registered speaker with the server and matching can be performed against voice matching devices connected to the server, making the process simple and convenient.

[0144]

[Effects of the Invention]

Because the invention "comprises: a plurality of voice inputting means for inputting voice signals of a speaker to be matched; a voice comparing means for comparing a plurality of voice signals inputted by each of the plurality of voice inputting means; a voice extracting means for extracting at least two of the approximate voice signals from the plurality of voice signals based on the results of the comparison performed by the voice comparing means; a voice correcting means for correcting one of the at least two approximate voice signals based on the at least two approximate voice signals extracted by the voice extracting means; and a voice recognizing means that performs a voiceprint and/or phonological analysis based on voice signals corrected by the voice correcting means, wherein the voice matching device has a function of matching voice signals from registered speakers with voice signals of the speaker to be matched based on the results of the voiceprint and/or phonological analysis performed by the voice recognizing means," correction is performed to remove noise that conceals differences by comparing a plurality of voice signals with those included in these voice signals. This makes it possible to collect high-quality voices with less influence from the surrounding environment than with voice signals inputted conventionally. This enables a voice matching device to be provided that has a higher chance of successful matching.

[Brief Description of the Drawings]

[FIG. 1]

FIG. 1 is a block diagram showing the electrical configuration of the voice matching device in an embodiment of the present invention.

[FIG. 2]

FIG. 2 is a diagram used to explain a database for the voice matching device in an embodiment of the present invention.

[FIG. 3]

FIG. 3 is a diagram used to explain a table for the voice matching device in an embodiment of the present invention.

[FIG. 4]

FIG. 4 is a block diagram showing the electrical configuration of a voice input device for the voice matching device in an embodiment of the present invention.

[FIG. 5]

FIG. 5 is a flowchart of operations performed by the voice matching device in an embodiment of the present invention.

[FIG. 6]

FIG. 6 is a flowchart of operations performed by the voice matching device in an embodiment of the present invention.

[FIG. 7]

FIG. 7 is a flowchart of operations performed by the voice matching device in an embodiment of the present invention.

[FIG. 8]

FIG. 8 is a flowchart of operations performed by the voice matching device in an embodiment of the present invention.

[FIG. 9]

FIG. 9 is a flowchart of operations performed by the voice matching device in an embodiment of the present invention.

[FIG. 10]

FIG. 10 is a flowchart of operations performed by the voice matching device in an embodiment of the present invention.

[FIG. 11]

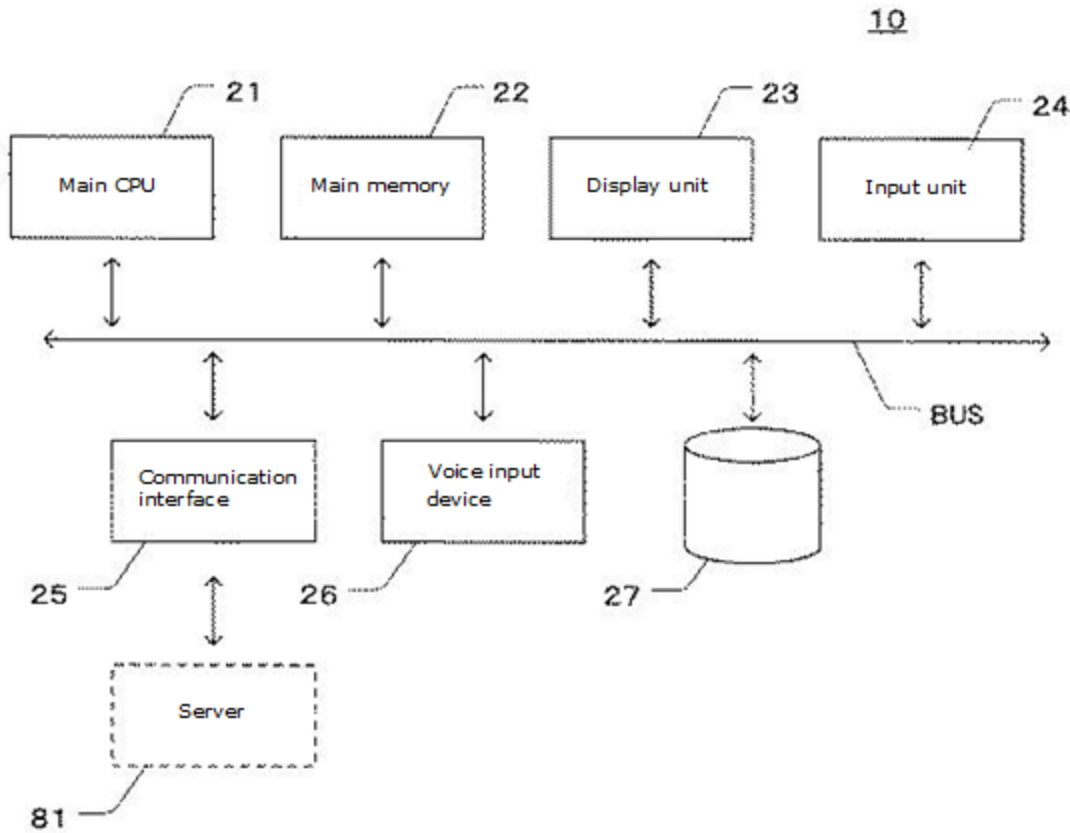
FIG. 11 is a block diagram showing the electrical configuration of a voice matching device including a server in an embodiment of the present invention.

[Reference Numbers]

10: Voice matching device
21: Main CPU
22: Main memory
23: Display unit
24: Input unit
25: Communication interface
26: Voice input device
27: Storage unit
52: CPU for voice input
54: ROM for voice input
56: RAM for voice input
58: Buffer
60: I/O bus
81: Server
102: Main microphone
104: Bandpass filter (BPF)
106: Low frequency amplifier
108: Sample and hold circuit
110: Converter
152: Voice data memory
202: Reference microphone
206: Low frequency amplifier
208: Sample and hold circuit
210: Converter
302: Cross correlation determiner
304: Noise canceling unit

306: Noise cancelled voice data memory
 308: Fast Fourier transform (FFT) processing unit
 310: Voice spectrum data memory

[FIG. 1]



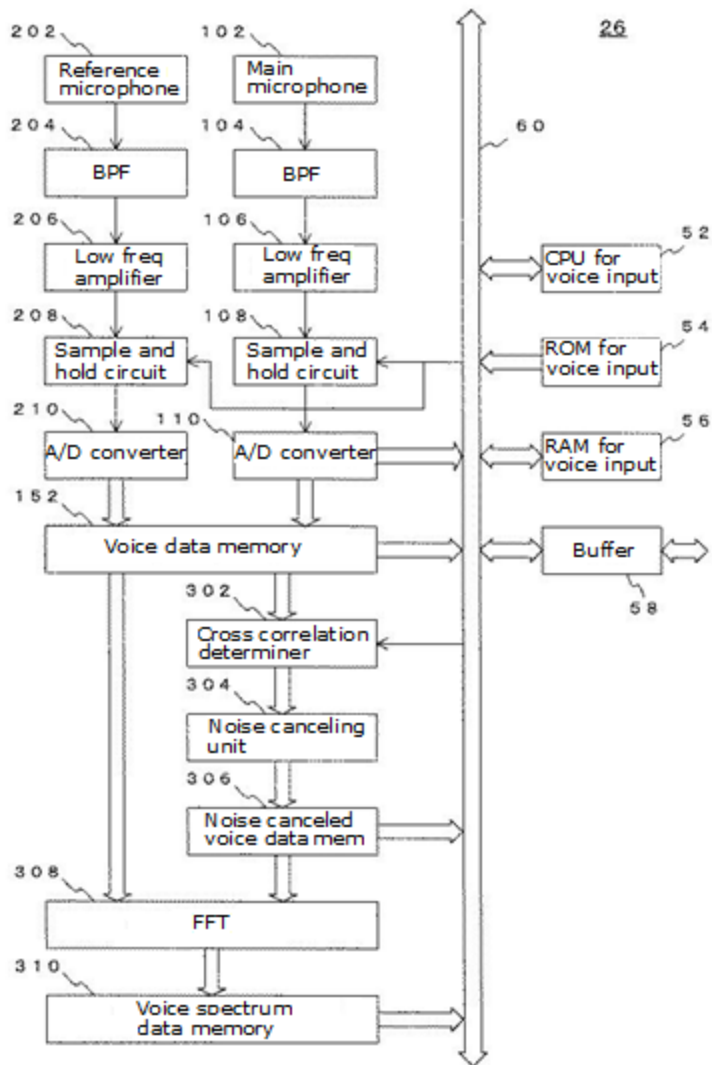
[FIG. 2]

ID	Question ID	Response Information	Voiceprint Information ID	...
0001	0001	It's Tokyo.	00010001	...
	0002	It's Yukinori.	00010002	...
	0003	It's Sumiko.	00010003	...
	0004	It's the Deccan Plateau.	00010004	...
	...	...	...	...
0002	0001	Gifu Prefecture	00020001	...
...	...	...	...	...

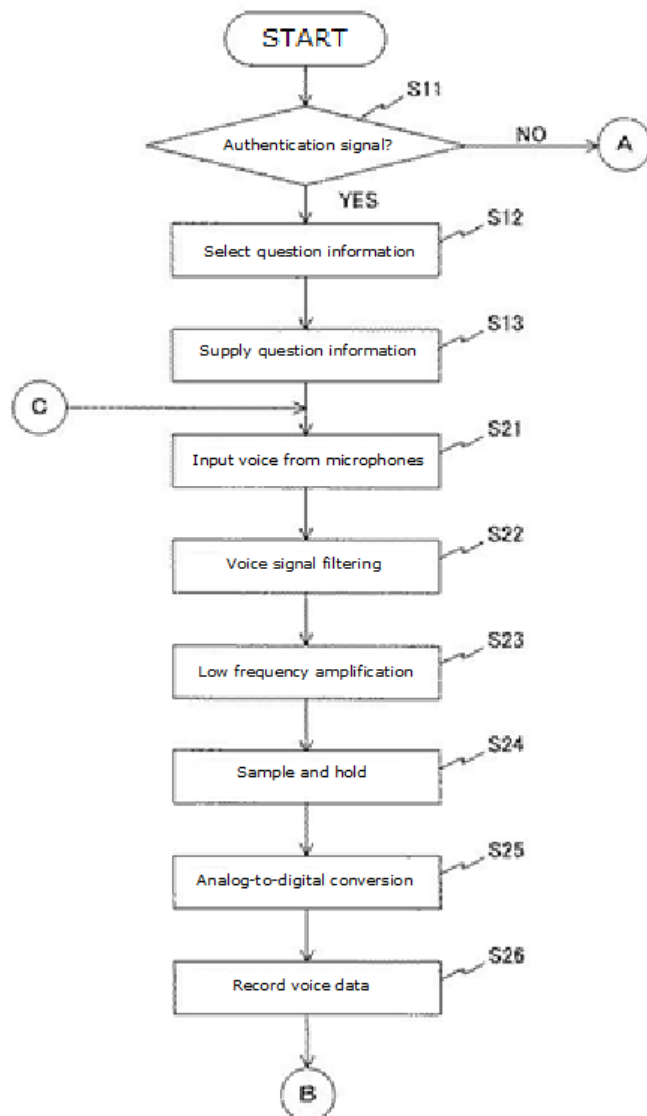
[FIG. 3]

Question ID	Question Information
0001	Where are you from?
0002	What is your father's name?
0003	What is your mother's name?
0004	Where would you like to travel?
...	...

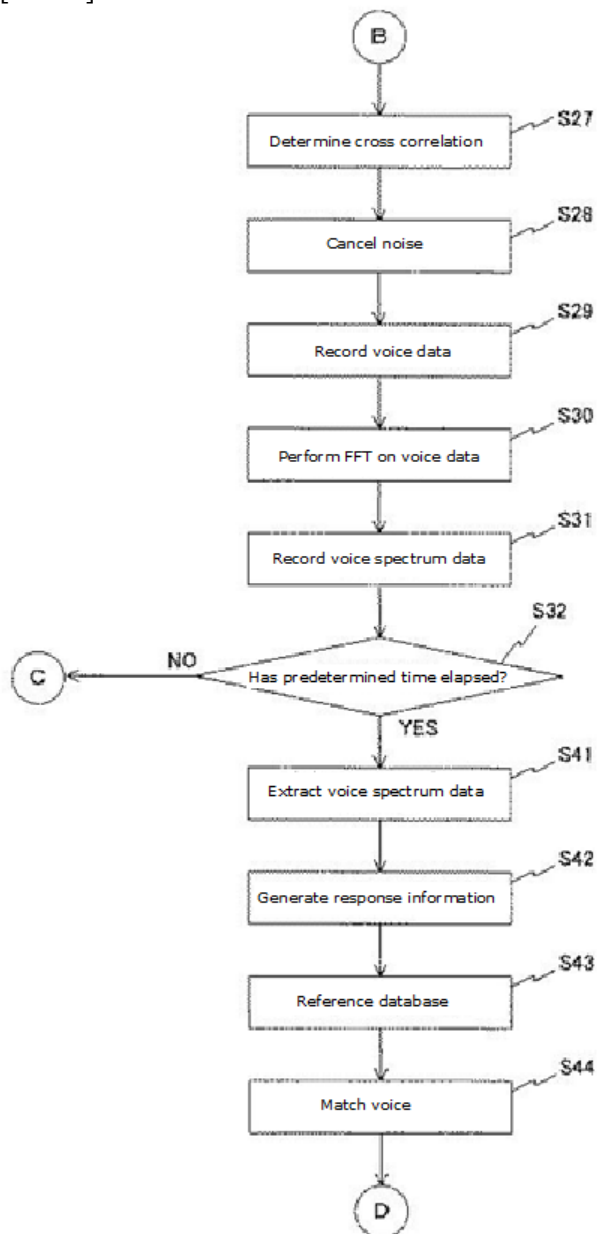
[FIG. 4]



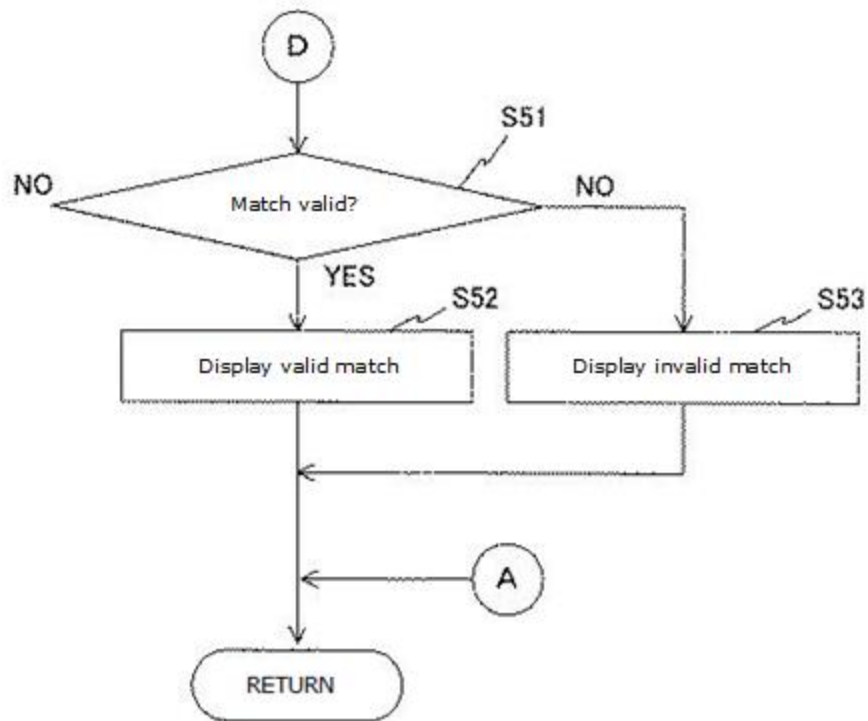
[FIG. 5]



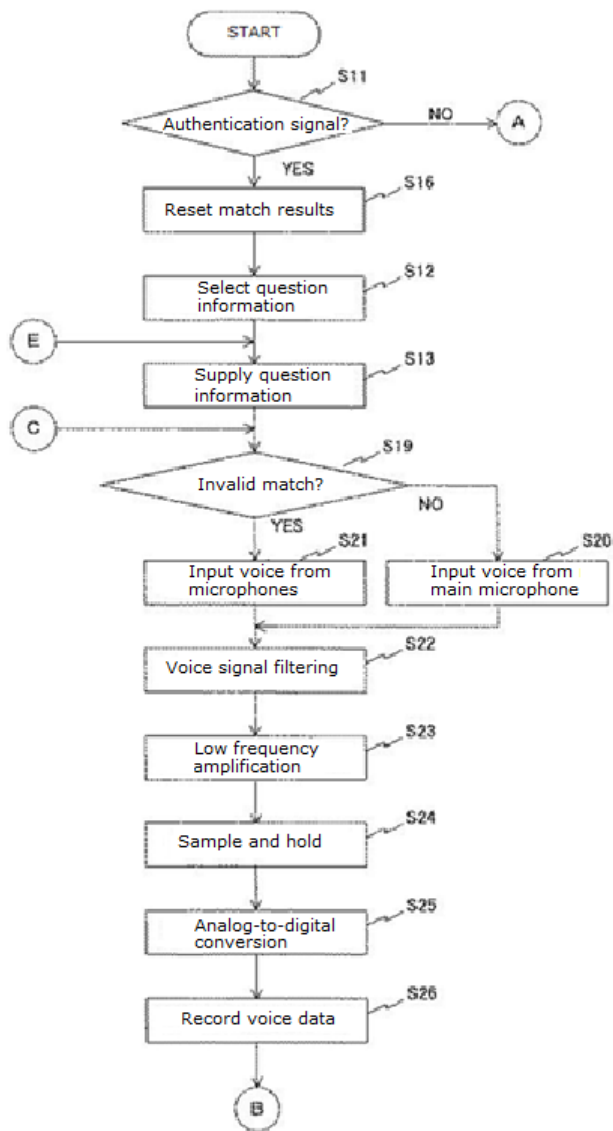
[FIG. 6]



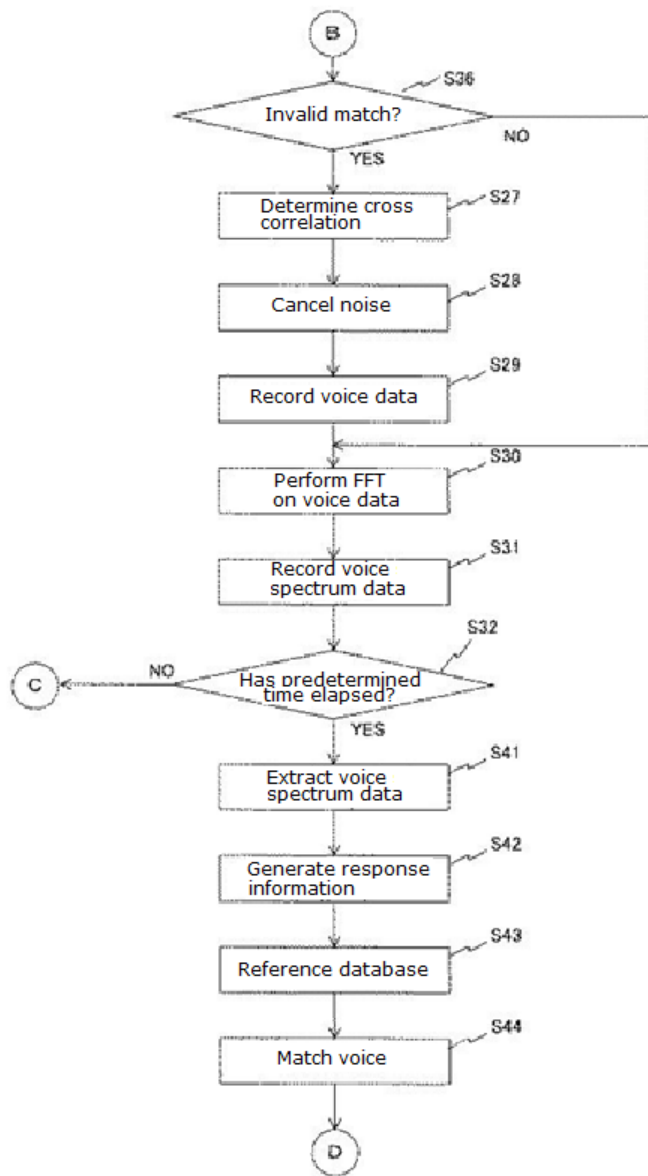
[FIG. 7]



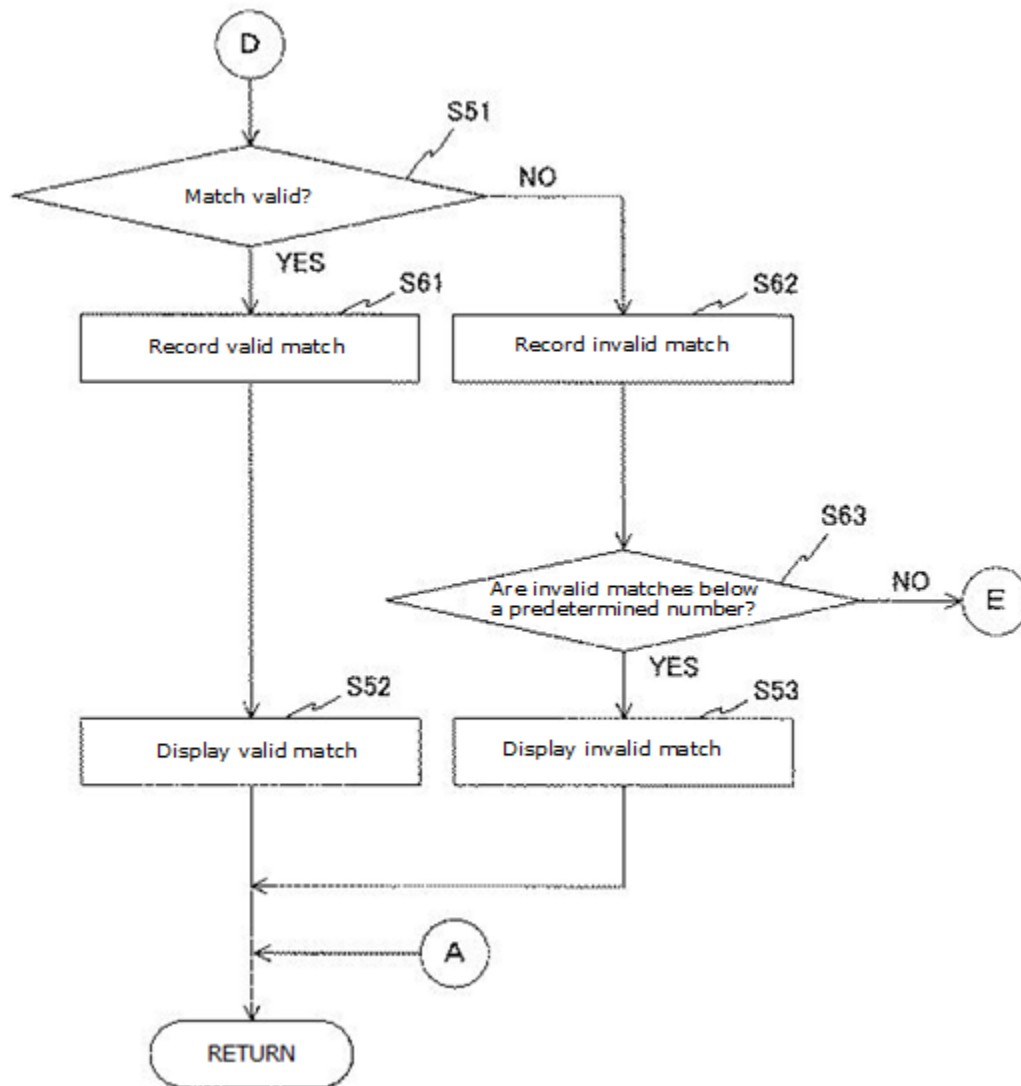
[FIG. 8]



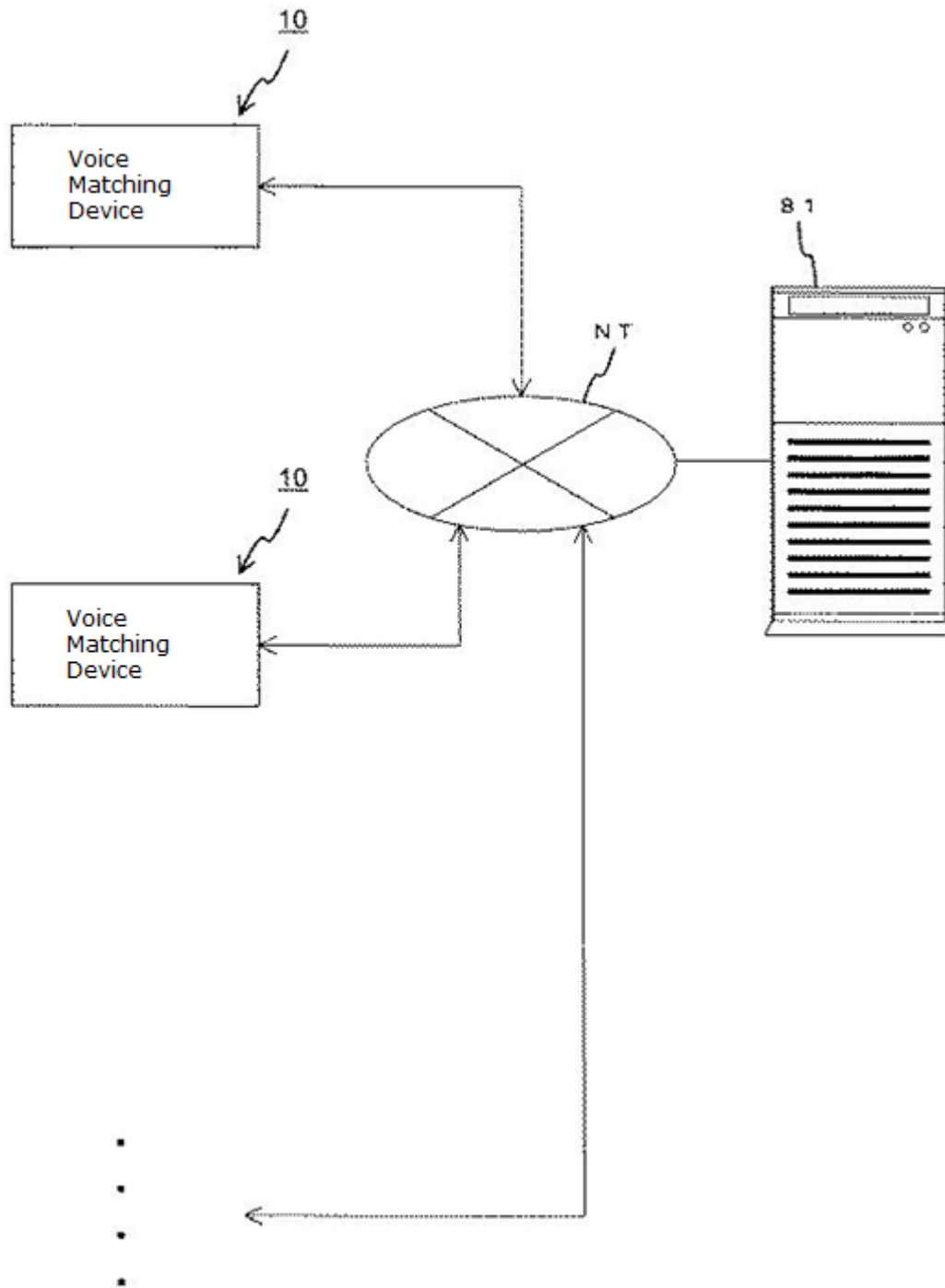
[FIG. 9]



[FIG. 10]



[FIG. 11]





Certification of Accuracy of Translation

Sun IP Project # 25-952

Japanese to English translation of JP2004145025A_ Original Fujimoto

I, Frank McGee, hereby certify that the attached English translation is, to the best of my knowledge, a true and accurate translation of “JP2004145025A_ Original Fujimoto”. The translated text reflects the content, meaning and style of the original text and constitutes a true and accurate translation of the original document.

March 4, 2025

A handwritten signature in blue ink, appearing to read "Frank McGee", written over a horizontal line.

Frank McGee