(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2019 (22.08.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/159253 A1

(51) International Patent Classification:
G10L 15/20 (2006.01) *G10L 15/28* (2013.01)

(21) International Application Number:
PCT/JP2018/005057

(22) International Filing Date:
14 February 2018 (14.02.2018)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **NEC CORPORATION** [JP/JP]; 7-1, Shiba 5-chome, Minato-ku, Tokyo, 1088001 (JP).

(72) Inventors: **WANG Qiongqiong**; c/o NEC Corporation, 7-1, Shiba 5-chome, Minato-ku, Tokyo, 1088001 (JP). **KOSHINAKA Takafumi**; c/o NEC Corporation, 7-1, Shiba 5-chome, Minato-ku, Tokyo, 1088001 (JP).

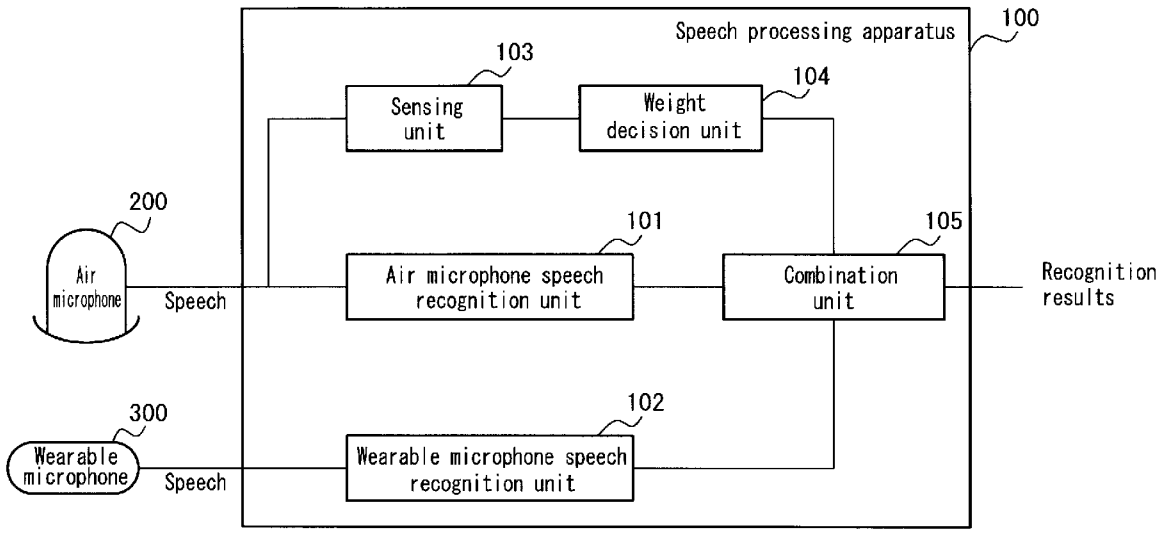
(74) Agent: **IWAKABE Fuyuki** et al.; SUNRISE PATENT OFFICE, Yomiuriyaesu Bldg. 6F, 8-7, Kyobashi 2-chome, Chuo-ku, Tokyo, 1040031 (JP).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) **Title:** SPEECH PROCESSING APPARATUS, METHOD, AND PROGRAM



(57) **Abstract:** The speech processing apparatus 100 includes an air microphone speech recognition unit 101 which recognizes speech from an air microphone 200 acquiring speech through air, a wearable microphone speech recognition unit 102 which recognizes speech from a wearable microphone 300, a sensing unit 103 which measures environmental conditions, a weight decision unit 104 which calculates the weights for recognition results of the air microphone speech recognition unit101 and the wearable microphone speech recognitionunit102 on the basis of the environmental conditions, and a combination unit 105which combines the recognition results outputted from the air microphone speech recognition unit101 and the wearable microphone speech recognition unit 102, using the weights.

WO 2019/159253 A1

Description

Title of Invention: SPEECH PROCESSING APPARATUS, METHOD, AND PROGRAM

Technical Field

[0001] The present invention relates to a speech processing apparatus, a speech processing method, and a speech processing program.

Background Art

[0002] The use of a The use of a microphone is the most popular method used to acquire voice in speech processing There are many types of microphones in accordance to different conduction methods, such as air microphones (air conduction microphones), throat microphones, bone microphones, skin microphones and so on. Among all of them, air microphones that acquire voice propagating through voice propagating through aiare the most commonly used in the speech processing field. An air microphone picks up the full-band, 20MHz for example, speech, by using which an appropriate speech processing system obtains good performance. However, it is less immune to environmental noise and far distance from the speech resource, referred to as speaker (refer to NPL1). How to improve the performance in such severe conditions is one of the most challenging topics among speech processing researchers. Especially now with the rapid spread of AI speakers such as Google Assistant, Amazon Alexa and Apple Siri, it has become a hotter and more urgent problem to solve (each of Google, Amazon, Alexa, Apple and Siri is a registered trademark).

[0003] A typical example of an air microphone is a sound collecting microphone which is used in a space such as a conference room, a hall or the like. However, usage of an air microphone is not limited to a conference, a convention and so on.

[0004] On the other hand, "wearable" devices have been increasing its popularity. Wearable devices are electronic devices for multiple purpose ranging from wireless transmission for communication objectives, music listening, fitness tracking, to biometric authentication. The hardware architecture of wearable devices typically comprises microphones. The microphone in a wearable device is referred to as "wearable microphone" in this specification. Generally, "wearable microphone" being less sensitive to surrounding noise and always on the speaker, i.e. close distance, may be used in noisy environments as mentioned above. However, since high-frequency components conducting through skin or bone attenuate significantly due to transmission loss, the quality of speech signals acquired by wearable microphone is relatively low. NPL 2 shows a wearable microphone performs worse than an air microphone in a speaker

recognition application in the clean condition.

Citation List

Patent Literature

[0005] PTL 1: Japanese Patent Application Laid-Open No. 2003-264883

Non Patent Literature

[0006] NPL 1: J. Ming, T. J. Hazen, J. R. Glass, and D. A. Reynolds, "Robust speaker recognition in noisy conditions", Audio, Speech, and Language Processing, IEEE Transactions, 2007

NPL 2: S. Tsuge, S. Kuroiwa, "Speaker recognition using multiple types of speech", The Seventh Symposium on Biometrics, Recognition and Authentication, 2017

NPL 3: E. Georgati, T. May, S. Par., et al. "Speaker Distance Detection Using a Single Microphone", IEEE TASLP 2011

NPL 4: J.G. Fiscus, "A post-processing system to yield reduced word error rates: Recognizer Output Voting Error Reduction (ROVER)", Automatic Speech Recognition and Understanding, 1997

Summary of Invention

Technical Problem

[0007] NPL1 which used only an air microphone showed that performance of speech processing was degraded by the effect of noise and the far distance between the air microphone and speakers which generates reverberation noise. It is proved that air microphones used in speech processing are vulnerable to far distance and noise. NPL2 compared in-ear bone microphone (a type of wearable microphones) and air microphone, and proved that in-ear bone microphone is not as competitive as the air microphone in speaker recognition in the clean and close condition.

[0008] In order to solve the above-mentioned problem, an aspect of the present invention is a speech processing apparatus using combination of different microphones by taking advantage of each type of microphones.

[0009] Additionally, PTL 1 discloses usage of different microphones, i.e. an air conduction sound microphone and a bone conduction sound microphone. However, the apparatus disclosed in PTL 1 adds a speech signal (a voice signal) outputted from the air conduction sound microphone to a speech signal outputted from the bone conduction sound microphone in accordance with the result of noise-level measurement. The apparatus does not take into account the effect of reverberation typically caused when the air conduction sound microphone is distant from the speaker. Addition of speech signals from different microphones, as done in the apparatus, does not necessarily improve the performance of recognition because such operations produce some signal which is quite different form original signals. It should be noted that recognition

systems are usually tuned or optimized to the microphone by which speech signals are captured.

Solution to Problem

[0010] An exemplary aspect of the speech processing apparatus includes air microphone speech recognition means for recognizing speech from an air microphone acquiring speech through air, wearable microphone speech recognition means for recognizing speech from a wearable microphone, sensing means for measuring environmental conditions including at least a distance between the air microphone and the speaker, weight decision means for calculating the weights for recognition results of the air microphone speech recognition means and the wearable microphone speech recognition means on the basis of the environmental conditions, and combination means for combining the recognition results outputted from the air microphone speech recognition means and the wearable microphone speech recognition means, using the weights.

[0011] An exemplary aspect of the speech processing method includes recognizing speech from an air microphone acquiring speech through air, recognizing speech from a wearable microphone, measuring environmental conditions including at least a distance between the air microphone and the speaker, calculating the weights for results of recognizing the speech from the air microphone and from the wearable microphone on the basis of the environmental conditions, and combining the results of recognizing the speech from the air microphone and from the wearable microphone, using the weights.

[0012] An exemplary aspect of the speech processing program causes a computer to execute recognizing speech from an air microphone acquiring speech through air, recognizing speech from a wearable microphone, measuring environmental conditions including at least a distance between the air microphone and the speaker, calculating the weights for results of recognizing the speech from the air microphone and from the wearable microphone on the basis of the environmental conditions, and combining the results of recognizing the speech from the air microphone and from the wearable microphone, using the weights.

Advantageous Effects of Invention

[0013] According to the present invention, the speech processing apparatus, speech processing method, and program of the present invention can provide better performance for speech processing in severe conditions by taking advantage of each type of microphones.

Brief Description of Drawings

[0014] [fig.1]Fig. 1 is a block diagram of a speech processing apparatus of the first example embodiment.

[fig.2]Fig. 2 is a flowchart illustrating an operation of the speech processing apparatus of the first example embodiment.

[fig.3]Fig. 3 is a block diagram illustrating a configuration example of an information processing system capable of implementing the functions of the speech processing apparatus.

[fig.4]Fig. 4 is a block diagram illustrating main parts of the speech processing apparatus.

Description of Embodiments

[0015] One example embodiment of the present invention will be described below with reference to the figures. The following detailed descriptions are merely exemplary in nature and are not intended to limit the invention or the application and uses of the invention. Furthermore, there is no intention to be bound by any theory presented in the preceding background of the invention or the following detailed description. In addition, the drawings together with the detailed description, serve to explain the principles for the inventive method. The drawings are for illustration and do not limit the application of the technique.

[0016] First Example embodiment.

A speech processing apparatus of the first example embodiment can provide high performance of speech processing in severe conditions of noise or far distance between the speaker and the air microphone.

[0017] < Configuration of the speech processing apparatus >

Fig. 1 illustrates a block diagram of a speech processing apparatus 100 of the first embodiment. The speech processing apparatus 100 includes an air microphone speech recognition unit (recognition unit using regular air microphone) 101, a wearable microphone speech recognition unit 102, a sensing unit 103, a weight decision unit 104, and a combination unit 105.

[0018] The air microphone speech recognition unit 101 applies speech processing for recognition using the speech data from an air microphone 200. In the case of speaker recognition which is to predict who is speaking, the air microphone speech recognition unit 101 uses the speech data from the air microphone 200 and calculates scores for the potential speakers. Log likelihood is often used as the score here. In case of speech recognition which is to predict the speech contents, the air microphone speech recognition unit 101 uses the speech from the air microphone 200 and calculates the

posteriors of linguistic units, e.g. phonemes, syllables, words, as the scores. Any kind of speech processing is possible: speech recognition,, emotion recognition, language recognition, dialect recognition, age recognition, gender recognition, etc. Details of all recognition are well-known for a person skilled in the art.

[0019] The wearable microphone speech recognition unit 102 applies speech processing for recognition using the speech data from a wearable microphone 300. An example of the wearable microphone 300 is an in-ear microphone. In the case of speaker recognition, the wearable microphone speech recognition unit 102 uses the speech data from the wearable microphone 300 and calculates scores for the potential speakers. In case of speech recognition, the wearable microphone speech recognition unit 102 uses the speech data from the wearable microphone 300 and calculates the posteriors of linguistic units as the scores.

[0020] The sensing unit 103 measures the environmental conditions, including signal-noise-ratio (SNR) and distance between the air microphone 200 and the speaker. For SNR measurement, the sensing unit 103 first detects speech and non-speech segments in the air microphone speech, and then calculates the ratio between average amplitudes of speech and non-speech segments. As for distance measurement, plural ways can be used here. For example, in-door positioning sensor, which is embedded together with the wearable microphone 300 and not shown in the figure, acquires the position of the speaker and calculates the distance between it and the known position of the air microphone 200. Another way is speaker distance detection using a single microphone as is disclosed in NPL 3.

[0021] The weight decision unit 104 decides weights according to the sensing result by the sensing unit 103. The weights here refer to the weights of using the results outputted from the air microphone speech recognition unit 101 and the wearable microphone speech recognition unit 102. For example, the weights w_a and w_s is calculated using following equations (1) and (2).

[0022] $w_a = a + c*SNR$ (1)

$w_s = b + d*distance$ (2)

[0023] Here, "distance" is a distance between the air microphone 200 and the speaker measured by the sensing unit 103. w_a is a weight for each score s_a from the air microphone speech recognition unit 101. w_s is a weight for each score s_s from the wearable microphone speech recognition unit 102. a, b, c and d are constants which can be determined experimentally. c and d are positive.

[0024] The example shown here is based on 1st-order polynomial functions in calculating the weights w_a and w_s . However, it is of course possible to use any other functions that are monotonically non-decreasing with respect to SNR and distance.

[0025] Another example of such functions is a step function that can take only two values of

0 or 1. In that case, either of the air microphone speech recognition unit 101 or the wearable microphone speech recognition unit 104 is exclusively chosen in accordance with SNR and distance.

[0026] The combination unit 105 combines the scores from the air microphone speech recognition unit 101 and the wearable microphone speech recognition unit 102, using the weights outputted from the weight decision unit 104, and output the final recognition result. In the case of speaker recognition, for each potential speakers, the combination unit 105 calculates a combined score s using following equation (3), for example.

[0027] $s = (w_a s_a + w_s s_s) / (w_a + w_s)$ (3)

[0028] By comparing the scores of all potential speakers, the combination unit 105 outputs information (data) indicating the speaker who has the highest score as the recognition result. In the case of speech recognition, Recognizer Output Voting Error Reduction (ROVER) in NPL4 can be used in the combination unit 105.

[0029] < Operation of the speech processing apparatus >

Next, the operation of the speech processing apparatus 100 will be described with reference to Fig. 2.

[0030] In step S11, the air microphone speech recognition unit 101 and the wearable microphone speech recognition unit 102 receive speech data.

[0031] In step S12, the air microphone speech recognition unit 101 estimates recognition results by using the speech data from the air microphone 200. Namely, the air microphone speech recognition unit 101 calculates score s_a, for example. In the case of speaker recognition, the air microphone speech recognition unit 101 calculates scores for the potential speakers. In the case of speech recognition, emotion recognition, language recognition, dialect recognition, age recognition, or gender recognition, the air microphone speech recognition unit 101 calculates scores for predetermined potential speech, potential emotion, potential words, potential dialect words, potential words for each age, or potential words for each gender.

[0032] In step S13, the wearable microphone speech recognition unit 102 estimates recognition results for example, by using the speech data from the wearable microphone 300 in the same way as the air microphone speech recognition unit 101. Namely, the wearable microphone speech recognition unit 102 calculates score s_s, for example.

[0033] In step S14, the sensing unit 103 measures environmental conditions including SNR and distance between the air microphone and the speaker.

[0034] In step S15, the weight decision unit 104 decides the weights for recognition results from the air microphone speech recognition unit 101 and the wearable microphone speech recognition unit 102.

[0035] Note that the order of steps A02-A03 and steps A04-A05 can be switched, not limited to the form presented in Fig. 2. Step S12 and step S13 can be also switched.

[0036] In step S16, the combination unit 105 combines the recognition results outputted from the air microphone recognition speech unit 101 and the wearable microphone speech recognition unit 102, while reflecting the weights outputted from the weight decision unit 104.

[0037] As explained above, the first example embodiment can improve performance of speech processing for recognition in severe conditions. It takes the advantages of both the air microphone 200 that picks up full-band speech and the wearable microphone 300 that is robust to surrounding noise. In other words, the speech processing apparatus, speech processing method, and program of the above example embodiment can provide better performance for speech processing in severe conditions by taking advantage of each type of microphones that the air microphone 200 picks up full-band speech and the wearable microphone 300 is robust to surrounding noise.

[0038] As a final point, it should be clear that the process, techniques and methodology described and illustrated here are not limited or related to a particular apparatus. It can be implemented using a combination of components. Also various types of general purpose devise may be used in accordance with the instructions herein. The present invention has also been described using a particular set of examples. However, these are merely illustrative and not restrictive. For example the described software may be implemented in a wide variety of languages such as C/C++, Java, MATLAB and Python etc. Moreover other implementations of the inventive technology will be apparent to those skilled in the art.

[0039] The aforementioned example embodiment can be configured in hardware, but it is also possible to implement an example embodiment by a computer program.

[0040] An information processing system shown in Fig. 3 includes one or more processors such as CPUs (Central Processing Unit) 1000, an I/O (Input/Output) device 1001 and a memory 1002. A magnetic storage medium such as a hard disk, or a semiconductor memory such as a ROM (read only memory) or a flash memory can be used as the memory 1002. The I/O device 1001 can realize functions of an output unit for inputting speech data and an input unit for outputting recognition results.

[0041] The memory 1002 stores a program for realizing the function of each unit in the speech processing apparatus 100 shown in Fig. 1. In the memory 1002, an area in which at least a program is stored is a non-transitory tangible storage area (non-transitory tangible media). In addition, the memory 1002 may realize a temporary information storage unit. The CPU 1000 executes processes according to the program stored in the memory 1002 to realize functions of the air microphone speech recognition unit 101, the wearable microphone speech recognition unit 102, the

sensing unit 103, the weight decision unit 104, and the combination unit 105 shown in Fig. 1.

[0042] Fig. 4 is a block diagram showing main parts of a speech processing apparatus according to the present invention. As shown in Fig. 4, the speech processing apparatus 10 includes air microphone speech recognition means 11 (realized by the air microphone speech recognition unit 101 in the example embodiment) for recognizing speech from an air microphone acquiring speech through air, wearable microphone speech recognition means 12 (realized by the wearable microphone speech recognition unit 102 in the example embodiment) for recognizing speech from a wearable microphone, sensing means13 (realized by the sensing unit 103 in the example embodiment) for measuring environmental conditions including at least a distance between the air microphone and the speaker, weight decision means 14 (realized by the weight decision unit 104 in the example embodiment) for calculating the weights for recognition results of the air microphone speech recognition means 11 and the wearable microphone speech recognition means 12 on the basis of the environmental conditions, and combination means 15 (realized by the combination unit 105 in the example embodiment) for combining the recognition results outputted from the air microphone speech recognition means 11 and the wearable microphone speech recognition means 12, using the weights.

[0043] The foregoing exemplary embodiments may be partly or wholly described in the following supplementary notes, though the structure of the present invention is not limited to such.

[0044] (Supplementary note 1) A speech processing apparatus comprising:

air microphone speech recognition means for recognizing speech from an air microphone acquiring speech through air,

wearable microphone speech recognition means for recognizing speech from a wearable microphone,

sensing means for measuring environmental conditions including SNR and a distance between the air microphone and the speaker,

weight decision means for calculating the weights for recognition results of the air microphone speech recognition means and the wearable microphone speech recognition means on the basis of the environmental conditions, and

combination means for combining the recognition results outputted from the air microphone speech recognition means and the wearable microphone speech recognition means, using the weights.

[0045] (Supplementary note 2) The speech processing apparatus according to Supplementary note 1,

wherein the weight decision means determines the weight for the recognition results

of the wearable microphone speech recognition means, in accordance with the distance between the air microphone and the speaker.

[0046] (Supplementary note 3) The speech processing apparatus according to Supplementary note 2,

wherein the weight for the recognition results of the air microphone speech recognition means is monotonically non-decreasing functions, with respect to the distance between the air microphone and the speaker.

[0047] (Supplementary note 4) The speech processing apparatus according to any one of Supplementary notes 1 to 3,

wherein the air microphone speech recognition means and the wearable microphone speech recognition means perform either of speech recognition, speaker recognition, language recognition, dialect recognition, emotion recognition, gender recognition, age recognition.

[0048] (Supplementary note 5) A speech processing method comprising:

recognizing speech from an air microphone acquiring speech through air,

recognizing speech from a wearable microphone,

measuring environmental conditions including SNR and a distance between the air microphone and the speaker,

calculating the weights for results of recognizing the speech from the air microphone and from the wearable microphone on the basis of the environmental conditions, and

combining the results of recognizing the speech from the air microphone and from the wearable microphone, using the weights.

[0049] (Supplementary note 6) The speech processing method according to Supplementary note 5,

wherein the weight for the results of recognizing the speech from the wearable microphone is determined in accordance with the distance between the air microphone and the speaker.

[0050] (Supplementary note 7) The speech processing method according to Supplementary note 6,

wherein the weight for the results of recognizing the speech from the air microphone is monotonically non-decreasing functions, with respect to the distance between the air microphone and the speaker.

[0051] (Supplementary note 8) A speech processing program for causing a computer to execute:

recognizing speech from an air microphone acquiring speech through air,

recognizing speech from a wearable microphone,

measuring environmental conditions including SNR and a distance between the air microphone and the speaker,

calculating the weights for results of recognizing the speech from the air microphone and from the wearable microphone on the basis of the environmental conditions, and combining the results of recognizing the speech from the air microphone and from the wearable microphone, using the weights.

[0052] (Supplementary note 9) The speech processing program according to Supplementary note 8,

wherein the weight for the results of recognizing the speech from the wearable microphone is determined in accordance with the distance between the air microphone and the speaker.

[0053] (Supplementary note 10) The speech processing program according to Supplementary note 9,

wherein the weight for the results of recognizing the speech from the air microphone is monotonically non-decreasing functions, with respect to the distance between the air microphone and the speaker.

[0054] (Supplementary note 11) A non-transitory computer readable information recording medium storing a speech processing program, when executed by a processor, performs,

recognizing speech from an air microphone acquiring speech through air,

recognizing speech from a wearable microphone,

measuring environmental conditions including at least a distance between the air microphone and the speaker,

calculating the weights for results of recognizing the speech from the air microphone and from the wearable microphone on the basis of the environmental conditions, and

combining the results of recognizing the speech from the air microphone and from the wearable microphone, using the weights.

[0055] (Supplementary note 12) The computer readable information recording medium according to Supplementary note 11,

wherein the weight for the results of recognizing the speech from the wearable microphone is determined in accordance with the distance between the air microphone and the speaker.

[0056] (Supplementary note 13) The computer readable information recording medium according to Supplementary note 12,

wherein the weight for the results of recognizing the speech from the air microphone is monotonically non-decreasing functions, with respect to the distance between the air microphone and the speaker.

[0057] While the present invention has been described with reference to the exemplary embodiments and examples, the present invention is not limited to these exemplary embodiments and examples. Various changes understandable by those skilled in the art can be made to the structures and details of the present invention within the scope of

the present invention.

Reference Signs List

[0058] 10 speech processing apparatus
11 air microphone speech recognition means
12 wearable microphone speech recognition means
13 sensing mean
14 weight decision means
15 combination means
100 speech processing apparatus
101 air microphone speech recognition unit
102 wearable microphone speech recognition unit
103 sensing unit
104 weight decision unit
105 combination unit
200 air microphone
300 wearable microphone
1000 CPU
1001 I/O device
1002 memory

Claims

[Claim 1] A speech processing apparatus comprising:
air microphone speech recognition means for recognizing speech from an air microphone acquiring speech through air,
wearable microphone speech recognition means for recognizing speech from a wearable microphone,
sensing means for measuring environmental conditions including at least a distance between the air microphone and the speaker,
weight decision means for calculating the weights for recognition results of the air microphone speech recognition means and the wearable microphone speech recognition means on the basis of the environmental conditions, and
combination means for combining the recognition results outputted from the air microphone speech recognition means and the wearable microphone speech recognition means, using the weights.

[Claim 2] The speech processing apparatus according to claim 1,
wherein the weight decision means determines the weight for the recognition results of the wearable microphone speech recognition means, in accordance with the distance between the air microphone and the speaker.

[Claim 3] The speech processing apparatus according to claim 2,
wherein the weight for the recognition results of the air microphone speech recognition means is monotonically non-decreasing functions, with respect to the distance between the air microphone and the speaker.

[Claim 4] The speech processing apparatus according to any one of claims 1 to 3,
wherein the air microphone speech recognition means and the wearable microphone speech recognition means perform either of speech recognition, speaker recognition, language recognition, dialect recognition, emotion recognition, gender recognition, age recognition.

[Claim 5] A speech processing method comprising:
recognizing speech from an air microphone acquiring speech through air,
recognizing speech from a wearable microphone,
measuring environmental conditions including at least a distance between the air microphone and the speaker,
calculating the weights for results of recognizing the speech from the

air microphone and from the wearable microphone on the basis of the environmental conditions, and

combining the results of recognizing the speech from the air microphone and from the wearable microphone, using the weights.

[Claim 6] The speech processing method according to claim 5,

wherein the weight for the results of recognizing the speech from the wearable microphone is determined in accordance with the distance between the air microphone and the speaker.

[Claim 7] A speech processing program for causing a computer to execute:

recognizing speech from an air microphone acquiring speech through air,

recognizing speech from a wearable microphone,

measuring environmental conditions including at least a distance between the air microphone and the speaker,

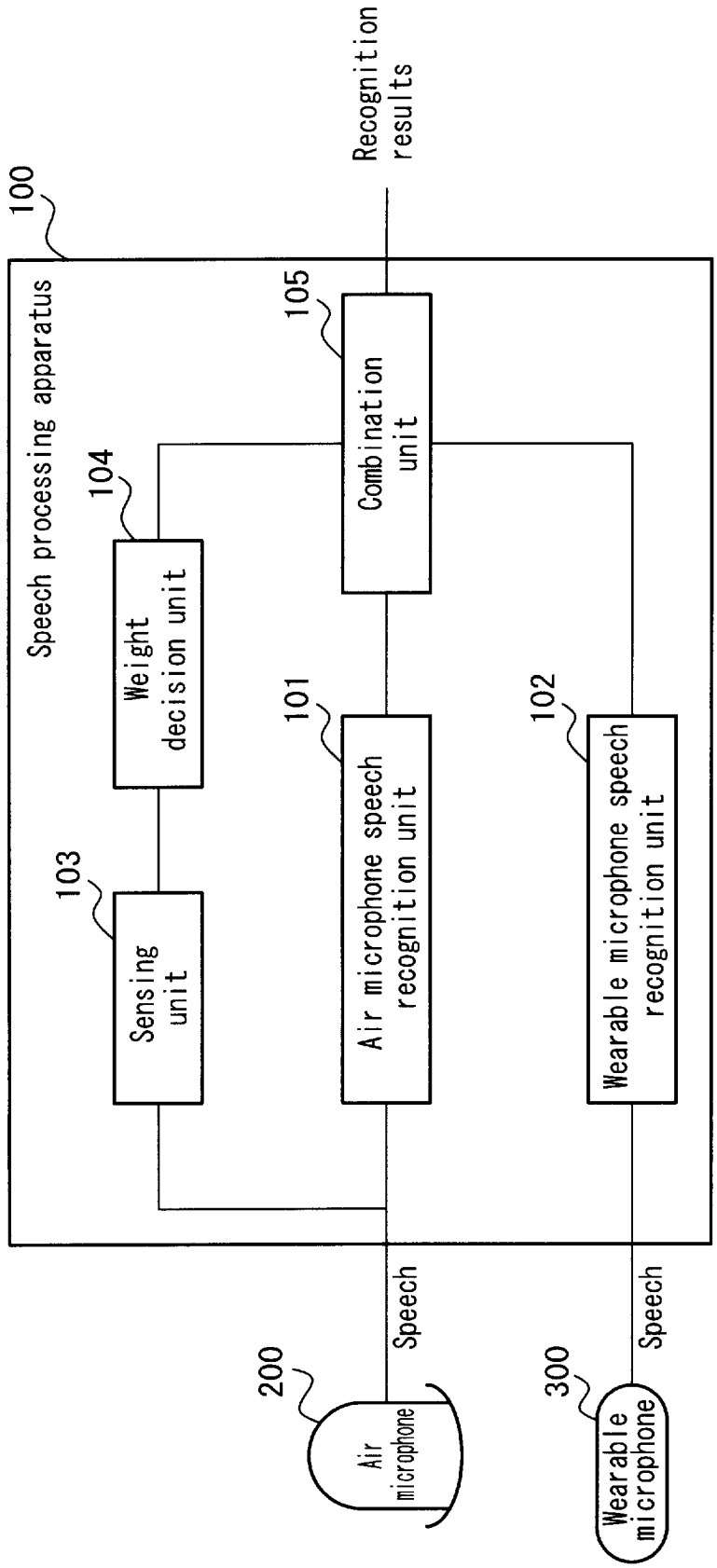
calculating the weights for results of recognizing the speech from the air microphone and from the wearable microphone on the basis of the environmental conditions, and

combining the results of recognizing the speech from the air microphone and from the wearable microphone, using the weights.

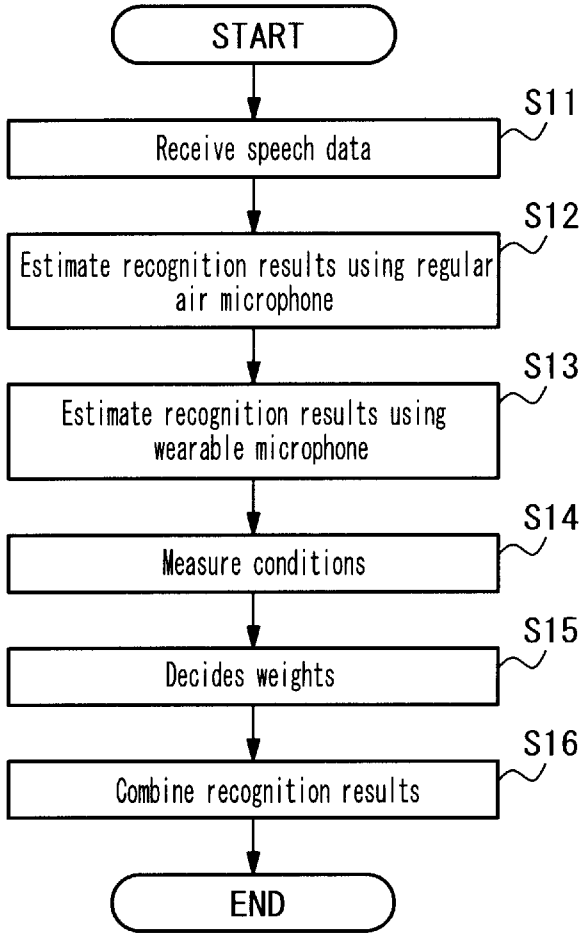
[Claim 8] The speech processing program according to claim 7,

wherein the weight for the results of recognizing the speech from the wearable microphone is determined in accordance with the distance between the air microphone and the speaker.

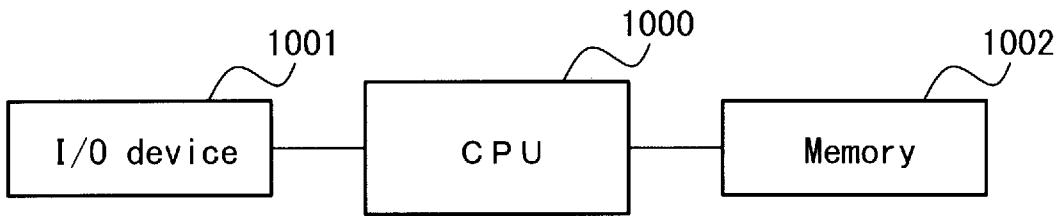
[Fig. 1]


Recognition results
100
Speech processing apparatus
105
Combination unit
104
Weight decision unit
103
Sensing unit
101
Air microphone speech recognition unit
102
Wearable microphone speech recognition unit
Speech
Speech
200
Air microphone
300
Wearable microphone

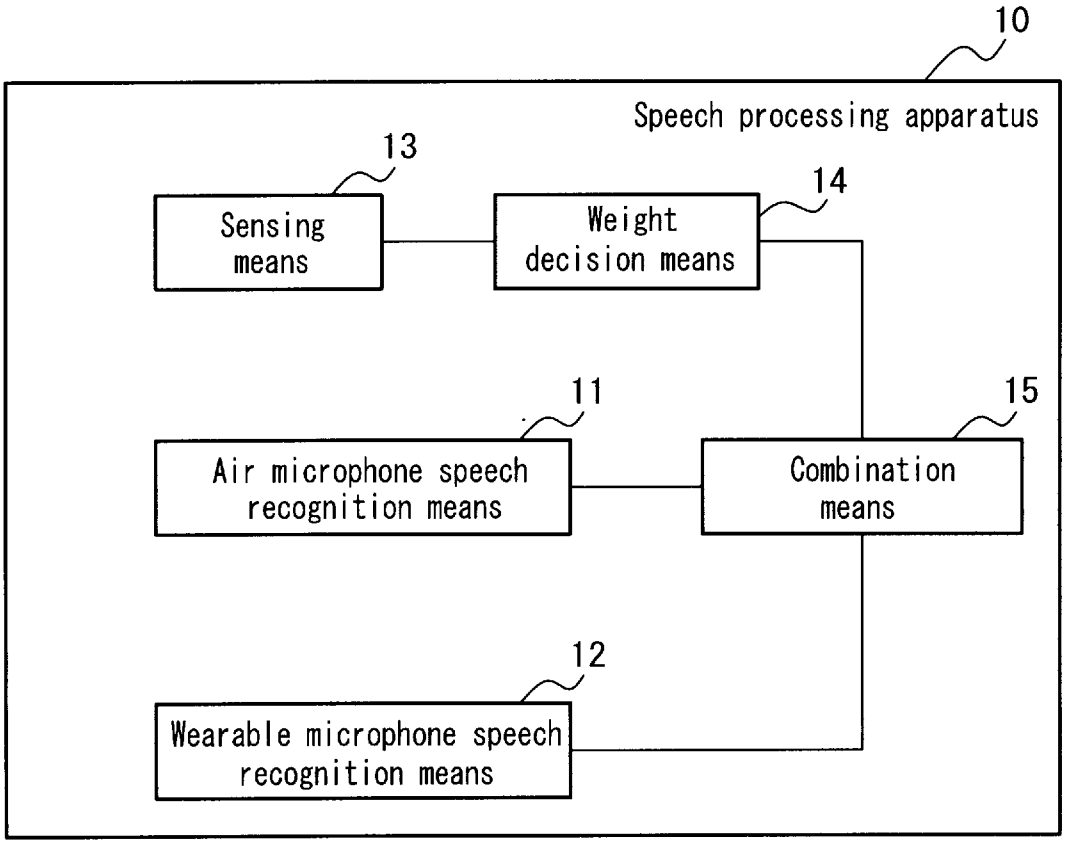
[Fig. 2]


START
S11
Receive speech data
S12
Estimate recognition results using regular air microphone
S13
Estimate recognition results using wearable microphone
S14
Measure conditions
S15
Decides weights
S16
Combine recognition results
END

[Fig. 3]


1001
1000
1002
I/O device
CPU
Memory

[Fig. 4]


10
Speech processing apparatus
13
Sensing means
14
Weight decision means
11
Air microphone speech recognition means
15
Combination means
12
Wearable microphone speech recognition means

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2018/005057

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G10L15/20(2006.01)i, G10L15/28(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl. G10L15/20, G10L15/28

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2018
Registered utility model specifications of Japan 1996-2018
Published registered utility model applications of Japan 1994-2018

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

IEEE Xplore

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TSUGE, Satoru et al., Speaker verification method using bone-conduction and air-conduction speech, 2009 International Symposium on Intelligent Signal Processing and Communication Systems(ISPACS2009), IEEE, 2009.12.09, pp.449-452, ISBN:978-1-4244-5016-9	1-8
A	JP 10-232691 A (NEC ROBOTICS ENG LTD.) 1998.09.02, paragraphs[0016]-[0017] (Family:none)	1-8
A	JP 2004-198831 A (SONY CORP.) 2004.07.15, paragraph[0106] (Family:none)	1-8
A	JP 11-305792 A (NIPPON TELEGRAPH & TELEPHONE) 1999.11.05, figs. 1-21 (Family:none)	1-8

☑ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24.04.2018	15.05.2018

Name and mailing address of the ISA/JP	Authorized officer
Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	TOMIZAWA, Naoki 5Z 4188 Telephone No. +81-3-3581-1101 Ext. 3591

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2018/005057

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2003-264883 A (DENSO CORP.) 2003.09.19, figs. 1-12 (Family:none)	1-8