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

(51) **International Patent Classification:**
G09B 19/04 (2006.01)

(21) **International Application Number:**
PCT/IB2007/051787

(22) **International Filing Date:** 11 May 2007 (11.05.2007)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
06114304.6 22 May 2006 (22.05.2006) EP

(71) **Applicant** *(for DE only)*: **PHILIPS INTELLECTUAL PROPERTY & STANDARDS GMBH** [DE/DE]; Lübeckertordamm 5, 20099 Hamburg (DE).

(71) **Applicant** *(for all designated States except DE, US)*: **KONINKLIJKE PHILIPS ELECTRONICS N.V.,** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **WILLMANN, Richard Daniel** [DE/NL]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). **LANFERMANN, Gerd** [DE/NL]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

GELLER, Dieter [DE/NL]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) **Agent: SCHOUTEN, Marcus M.**; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

[Continued on next page]

(54) **Title:** SYSTEM AND METHOD OF TRAINING A DYSARTHRIC SPEAKER
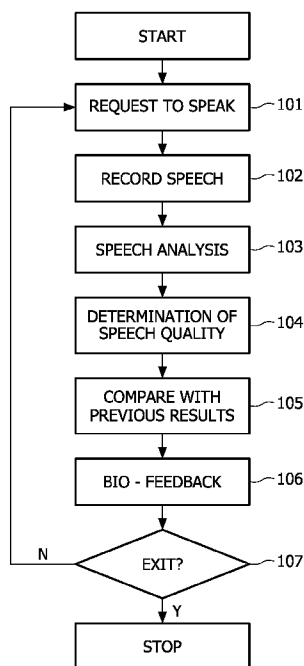
(57) **Abstract:** The invention describes a computer-based system that asks (101) a patient to pronounce a word displayed on a monitor, automatically assesses (104, 105) the speech quality, and uses suitable means to feed back (106) any improvement or deterioration of speech quality.

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

System and method of training a dysarthric speaker

5

The present invention relates to a system and method of training a dysarthric speaker. Furthermore, the invention relates to a computer program for training a dysarthric speaker.

10   Speech deficits are a common symptom of neurological diseases such as stroke. In the US, 700,000 new cases of stroke occur per year. While deadly in 20% of the cases, stroke causes impairments in over 60% of the victims. About 25% of the patients show speech deficits. Rehabilitation measures and loss of work force inflict costs of $50B to the US society per year.

15   There are two large groups of speech deficits: aphasia and dysarthria. Aphasia is an inability of the patient to name objects correctly. Dysarthria is the inability of the patient to pronounce words clearly and correctly.

Speech therapists exercise with dysarthric speakers in one-to-one sessions, usually by having the patient repeat a word. The patient receives no detailed corrections.

20   Obviously, this form of therapy is cost intensive and very demanding on the therapist.

To maximize rehabilitation efforts while optimizing costs, it is desirable that rehabilitation exercises are continued by the patient at home without a supervising professional being available. It is well known that appropriate feedback enhances rehabilitation success. To this end, automatic assessment of exercises is necessary, e.g. by means of a computer.

25   Aphasia is relatively easy to test for on a computer, e.g. by displaying an image of an object and asking the patient in a multiple choice fashion what the correct name of the object is. Training software is commercially available and used in rehabilitation practice. However, dysarthria is harder to test for on a computer, due to the potentially severe speech condition. Presently, systems for unsupervised training of impaired speech are only available

30   for aphasia, not for dysarthria.

It is an object of the present invention to provide an inexpensive, reliable and

unsupervised technique of training a dysarthric speaker.

This object is achieved according to the invention by a system for training a dysarthric speaker, the system comprising a first computing module adapted to automatically determine a speech quality measure for an input utterance of the speaker, a second computing module adapted to automatically compare the speech quality measure with at least one previous speech quality measure of the speaker, and a feedback module adapted to automatically provide a biofeedback to the speaker based upon this comparison.

The object of the present invention is also achieved by a method of training a dysarthric speaker, comprising the steps of automatically determining a speech quality measure for an input utterance of the speaker, automatically comparing the speech quality measure with at least one previous speech quality measure of the speaker, and based upon this comparison, automatically providing biofeedback to the speaker.

The object of the present invention is also achieved by a computer program to be executed in a computer, said program comprising computer instructions to automatically determine a speech quality measure for an input utterance of the speaker, computer instructions to automatically compare the speech quality measure with at least one previous speech quality measure of the speaker, and computer instructions to automatically provide biofeedback to the speaker, based upon this comparison, when the computer program is executed in the computer. The technical effects necessary according to the invention can thus be realized on the basis of the instructions of the computer program in accordance with the invention.

Such a computer program can be stored on a carrier such as a CD-ROM or it can be available over the Internet or another computer network. Prior to executing the computer program, it is loaded into the computer by reading the computer program from the carrier, for example by means of a CD-ROM player, or from the Internet, and storing it in the memory of the computer. The computer includes inter alia a central processor unit (CPU), a bus system, memory means, e.g. RAM or ROM etc., storage means, e.g. floppy disk or hard disk units etc. and input/output units. Alternatively, the inventive method could be implemented in hardware, e.g. using one or more integrated circuits.

A core idea of the invention is to provide a technique for automatically training a dysarthric speaker by means of a simple inexpensive training system. This idea can be realized for example by means of a standard computer system, a microphone coupled to the computer system and a dedicated computer program to be executed on the computer system.

Another core idea of the invention is that the speaker, e.g. the rehabilitating patient, improves his condition through the effect of biofeedback. Biofeedback is a technique in which a person is trained to improve his body condition by using signals from his own body. In the present case, information about the speech ability of the speaker is given back to the

5   speaker. Preferably, the feedback information depends on improvement or deterioration of speech quality. Using this feedback information, the speaker can take steps to improve his speech ability. The biofeedback mechanism allows the speaker to receive, e.g. to see or hear, information about his speech ability. For this purpose, an utterance of the speaker is "translated" into a signal, which can be detected by the speaker, e.g. a flashing light bulb, a

10   beeper or the like. For example, the improvement in speech quality compared to previous results is illustrated in form of a "traffic light" or specific sound. The speaker attempts to improve his skill and monitors his own performance. According to the results provided through biofeedback signals, the speaker "automatically" adjusts his speech. By using such a biofeedback mechanism, there is no need for detailed linguistic corrections. This reduces the

15   cost and effort of training considerably. Because the training is provided completely automatically, no treatment or training errors can occur. Thus, the training technique according to the present invention is very reliable.

According to the invention, the speaker is not trained to produce speech which is only understood by a machine. This aim could easily be reached by requesting the speaker to

20   repeat a certain word always in the same way. Instead, the system and method according to the present invention aim at improving the intelligibility of the users speech for every average human listener. For this reason, in a preferred embodiment of the invention, the measure of speech quality correlates to the deviation of the produced speech from an averaged normal speech production.

25   The present invention presents a realization of a technique for training speech quality of dysarthric speakers. It enables the user to train pronunciation without the presence of a speech therapist. Feedback is given to the patient that guides him towards a better speech quality.

With the present invention, an unsupervised training can be provided. Thus, the

30   invention can shorten or even eliminate a cost-intensive therapy including expensive sessions with speech therapists. To summarize, the present invention provides a simple, inexpensive and reliable technique for training a dysarthric speaker without the need of supervision.

These and other aspects of the invention will be further elaborated on the basis

4

of the following embodiments which are defined in the dependent claims.

According to a preferred embodiment of the invention, the system comprises a recording devise adapted to record the input utterance of the speaker. According to the inventive method, the input utterance of the speaker is recorded prior to determining the speech quality measure. The recording device, for example a microphone, is preferably connected to the first computing module in order to directly feed in the speaker's voice. Alternatively, the speaker's utterance is transmitted from the microphone to a pre-processing device, e.g. a voice filter or the like, which is adapted to pre-process the speaker's voice prior to the computing step carried out by means of the first computing module. Preferably, the speaker's utterance is processed and evaluated immediately after the utterance is made, i.e. the utterance is not stored in a database or the like. Thus, the training system can be designed in a very simple and inexpensive way.

According to another preferred embodiment of the invention, the system comprises an interface module adapted to provide a speaking instruction to the speaker. According to the inventive method, a speaking instruction is provided to the speaker prior to recording the input utterance of the speaker. For this purpose, the interface module of the training system preferably comprises a monitor means, e.g. a built-in display, for displaying the words to be repeated by the speaker. In other words, the speaker reads the words to be repeated. Instead of built-in monitor means, the interface module may be connected to an external display. Alternatively or additionally to the monitoring means, the interface module comprises a speech output device, e.g. a built-in loudspeaker, for playing back the words to be repeated by the speaker. In other words, the speaker listens to the word or words to be repeated. If such a speaking instruction is implemented, the speech analysis carried out by the first computing module can be much easier, faster and less complex, than in the case that the speaker speaks a random word. Thus, the first computing module can be built in a less complex manner, which leads to a significant cost reduction.

After the speaker's utterance is input into the training system, the system automatically assesses the speech quality. For this purpose, speech recognition technology is used.

According to another preferred embodiment of the invention, the first computing module comprises a speech analysis component adapted to carry out a speech recognition and speech analysis on a phoneme and/or word level. According to the inventive method, the step of determining the speech quality measure comprises carrying out a speech

analysis on a phoneme and/or word level. A speech analysis on a phoneme and/or word level is preferred because dysarthric speakers have in most cases deficits in the production of specific sounds. These sounds form the phonemes, the smallest units of speech that form the differences in the sounds of syllables, words and whole sentences. Speech analysis on a word

5   level allows the comparison of phoneme quality in similar words: e.g. 'cheese' and 'freeze' both contain the 'e:' phoneme. It is also possible to analyze phonemes in the context of neighboring ones: e.g. if a patient shows problems in pronouncing 'killed' while being able to speak 'cold', 'mold' and 'kitten', it becomes apparent that the sequence of 'i' and 'l' causes problems. Thus, speech analysis on a phoneme basis allows analyzing which specific sounds

10   deviate from the same sounds uttered by a plurality of non-dysarthric speakers.

        According to another preferred embodiment of the invention, the first computing module comprises a comparing component adapted to compare the input utterance with reference data, which is based upon data of a plurality of non-dysarthric speakers. According to the inventive method, the step of determining the speech quality measure

15   comprises comparing the input utterance with such reference data. Preferably, the reference data is based upon averaged data of a plurality of non-dysarthric speakers. For this purpose, data on a reference pronunciation is provided, e.g. in a database, for each word to be used in training. The reference pronunciation data has been collected prior to the training and represents the averaged pronunciation of words by a plurality of non-dysarthric speakers.

20   Based on the speech analysis and the comparison with reference data the first computing module computes an objective speech quality measure. The speech quality measure is preferably determined for each single training word. All results, i.e. speech quality measures for training words, are stored by the first computing module in a training database. Preferably the training results are stored together with a time stamp. By this means each result can later

25   by assigned to a specific date and time of day.

        According to another preferred embodiment of the invention, the second computing module comprises an evaluation component adapted to carry out a comparative evaluation of the speech quality measure with respect to previous results of the speaker. According to the inventive method, the step of comparing the speech quality measure

30   comprises carrying out a comparative evaluation of the speech quality measure with respect to previous results of the speaker. For this purpose, historic training data, i.e. a former speech quality measure for the respective word, is used for carrying out the comparison. Such historic training data is stored in the training database, which can be accessed by the second

6

computing module. Preferably, the second computing module evaluates the speech quality depending not only on previous results, but also on the time of day the training is performed. Thus, time-dependent performance fluctuations can be minimized or eliminated.

As a result of this evaluation of the speech quality measure with respect to previous results, information about improvement or impairment of pronunciation is provided, preferably for each single word. In other words, personal improvements are being used for biofeedback.

These and other aspects of the invention will be described in detail hereinafter, by way of example, with reference to the following embodiments and the accompanying drawings, in which:

Fig. 1 shows a schematic block diagram of a system according to the invention,

Fig. 2 shows a simplified flowchart of the method according to the invention.

In the described embodiment, the system 1 for training a dysarthric speaker is realized by means of a standard computer system 2, e.g. a notebook or PC, a microphone 3 coupled to the computer system 2 and a dedicated computer program to be executed on the computer system 2. The computer program executed on the computer system 2 realizes a first computing module 4 and a second computing module 5. Furthermore, the computer system 2 comprises a combined interface and feedback module 6. All appliances of the system 1 are adapted to carry out the method according to the present invention. All modules, devices and components are constructed and programmed in such a way that the procedures for data and/or information input, processing, and output run in accordance with the method of the invention. In particular the processing unit of the computer system 2 is adapted to perform all the tasks of calculating and computing the input data as well as determining and assessing results. This is achieved, according to the invention, by means of the dedicated computer software comprising computer instructions adapted for carrying out the steps of the inventive method, when the software is executed in the processing unit of the computer system 1. The processing unit itself may comprise functional modules or units, which are implemented in form of hardware, software or in form of a combination of both.

In a first step 101, a user (speaker) is requested to speak a word or short

sentence. For this purpose, the combined interface and feedback module 6 provides a speaking instruction to the speaker. The combined interface and feedback module 6 comprises the built-in monitor 7 of the computer system 2 or an external monitoring means, e.g. a television set, which is connected to the computer system. The combined interface and feedback module 6

5   further comprises a built-in loudspeaker 8 for playing back the words to be repeated by the speaker. For providing the speaking instruction the combined interface and feedback module 6 is controlled by the dedicated computer program.

In a next step 102, the user's speech, i.e. an input utterance of the speaker, is recorded by means of the microphone 3, which is connected to the computer system 2 and

10   adapted to transfer the speech signals to the computer system 2 for processing.

Subsequently, a speech quality measure for the input utterance of the speaker is computed by means of the first computing module 4. For this purpose, a speech recognition is carried out in step 103 by a speech analysis component 11 of the first computing module 4. The speech recognition is carried out on a phoneme and/or word level in order to determine

15   the spoken word or words. Subsequently, a comparing component 12 of the first computing module 4 compares in step 104 the input utterance with reference data, which is based upon averaged data of a plurality of non-dysarthric speakers. For this purpose, the comparing component 12 accesses a reference database 9, in which reference data for each training word is stored. Based on the speech analysis and the comparison with reference data, the first

20   computing module 4 computes an objective speech quality measure for each single training word.

Preferably, the first computing module 4 operates as follows: A word W is given for pronunciation training. The speaker's speech production may be a word O. The first computing module 4 now computes two probabilities. First, the probability that the given

25   speech production O matches the given word W, so-called "forced alignment". Second, the maximum of the probabilities that the given speech production O is an arbitrary series of phonemes, so called "free phoneme loop". The ratio R of the two probabilities (R=W/O) yields a measure for the pronunciation quality of the given word W. Details illustrating this approach can be found in the paper by H. Boulard and B. D'hoors 'Optimizing recognition

30   and rejection performance in word spotting systems', Proc. ICASSP, Vol 1, Page 373, April 1994.

Subsequently, the speech quality measures R are stored by the first computing module 4 in a training database 10 together with a time stamp and the corresponding word or

8

words.

In a next step 105, an evaluation component 13 of the second computing module 5 automatically compares the speech quality measure with at least one previous speech quality measure of the speaker. During this step, previous results may be taken into account by averaging all values of the current session and/or by using min/max values of the session and/or by weighting each value with regard to time (most recent values are more heavily weighted) followed by averaging and/or by taking into account accumulated results of previous sessions. For this purpose, the evaluation component accesses the training database 10, in which previous results for each training word are stored. If the system is used for the first time, the measured speech quality on a standard set of words is used for initializing the database 10 of historic training data and initializing the performance level relative to which biofeedback is given.

In a next step 106, a feedback is given to the speaker, which is based on the current exercise results. For this purpose, the combined interface and feedback module 6 provides a "traffic light" signaling scheme, in which in case of an improvement a green light is shown, in case of no improvement a yellow light is shown and in case of a deterioration a red light is shown to the speaker through the built-in monitor 7 of the computer system 2. For providing the feedback information, the combined interface and feedback module 6 is controlled by the computer program.

Preferably, the speaker is requested to repeat the training process until an exit condition is met (step 107). Such an exit condition may be for example: maximum training time reached, maximum "green" results in a row reached, maximum "yellow" results in a row reached, maximum "red" results in a row reached, maximum number of exercise runs exceeded, manual stop.

The general training results as well as the training results related to every single word or phrase are stored in a result database (not shown) and/or made available to the therapist by means of the computer system 2, e.g. by online data transfer to the therapist through a computer network, like the internet.

The present invention can preferably be used as part of the home stroke test, as part of a home-based rehabilitation application or in a rehabilitation clinic. The invention can further be used as part of a telemedicine platform.

It will be evident to those skilled in the art that the invention is not limited to the details of the foregoing illustrative embodiments, and that the present invention may be

embodied in other specific forms without departing from the spirit or essential attributes thereof. The present embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range

5 of equivalency of the claims are therefore intended to be embraced therein. It will furthermore be evident that the word "comprising" does not exclude other elements or steps, that the words "a" or "an" do not exclude a plurality, and that a single element, such as a computer system or another unit, may fulfil the functions of several means recited in the claims. Any reference signs in the claims shall not be construed as limiting the claim concerned.

10

REFERENCE NUMERALS:

	1	system
	2	computer
5	3	microphone
	4	first computing module
	5	second computing module
	6	combined interface and feedback module
	7	monitor
10	8	loudspeaker
	9	reference database
	10	training database
	11	speech analysis component
	12	comparing component
15	13	evaluation component

CLAIMS:

5

1.   A system (1) for training a dysarthric speaker, comprising

-        a first computing module (4) adapted to automatically determine a speech quality measure for an input utterance of the speaker,

-        a second computing module (5) adapted to automatically compare the speech

10   quality measure with at least one previous speech quality measure of the speaker, and

-        a feedback module (6) adapted to automatically provide a biofeedback to the speaker based upon this comparison.

2.   The system (1) as claimed in claim 1, comprising a recording device (3)

15   adapted to record the input utterance of the speaker.

3.   The system (1) as claimed in claim 1, comprising an interface module (6) adapted to provide a speaking instruction to the speaker.

20   4.   The system (1) as claimed in claim 1, wherein the first computing module (4) comprises a speech analysis component (11) adapted to carry out a speech analysis on a phoneme and/or word level.

5.   The system (1) as claimed in claim 1, wherein the first computing module (4)

25   comprises a comparing component (12) adapted to compare the input utterance with reference data, said reference data being based upon data of a plurality of non-dysarthric speakers.

6.   The system (1) as claimed in claim 1, wherein the second computing module (5) comprises an evaluation component (13) adapted to carry out a comparative evaluation of

30   the speech quality measure with respect to previous results of the speaker.

7.   A method of training a dysarthric speaker, comprising the steps of

-        automatically determining (104) a speech quality measure for an input utterance

of the speaker,

- automatically comparing (105) the speech quality measure with at least one previous speech quality measure of the speaker, and

- based upon this comparison, automatically providing (106) a biofeedback to the

5 speaker.

8. A computer program to be executed in a computer (2), said program comprising

- computer instructions to automatically determine (104) a speech quality

10 measure for an input utterance of the speaker,

- computer instructions to automatically compare (105) the speech quality measure with at least one previous speech quality measure of the speaker, and

- computer instructions to automatically provide (106) a biofeedback to the speaker, based upon this comparison,

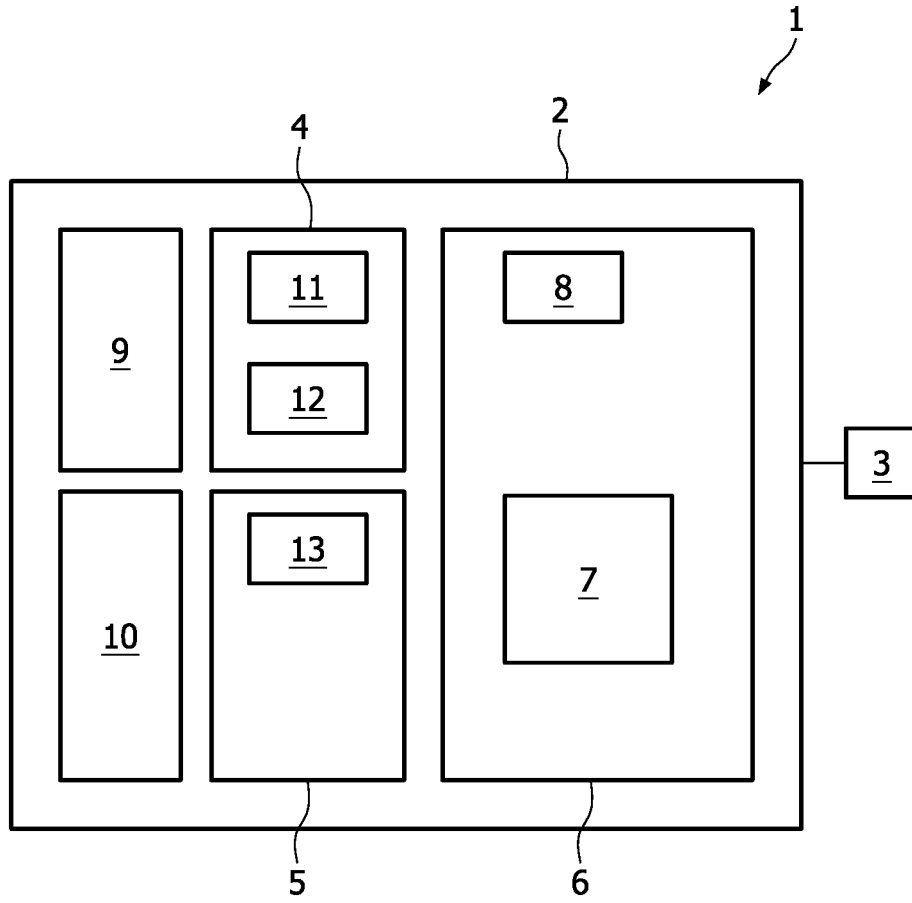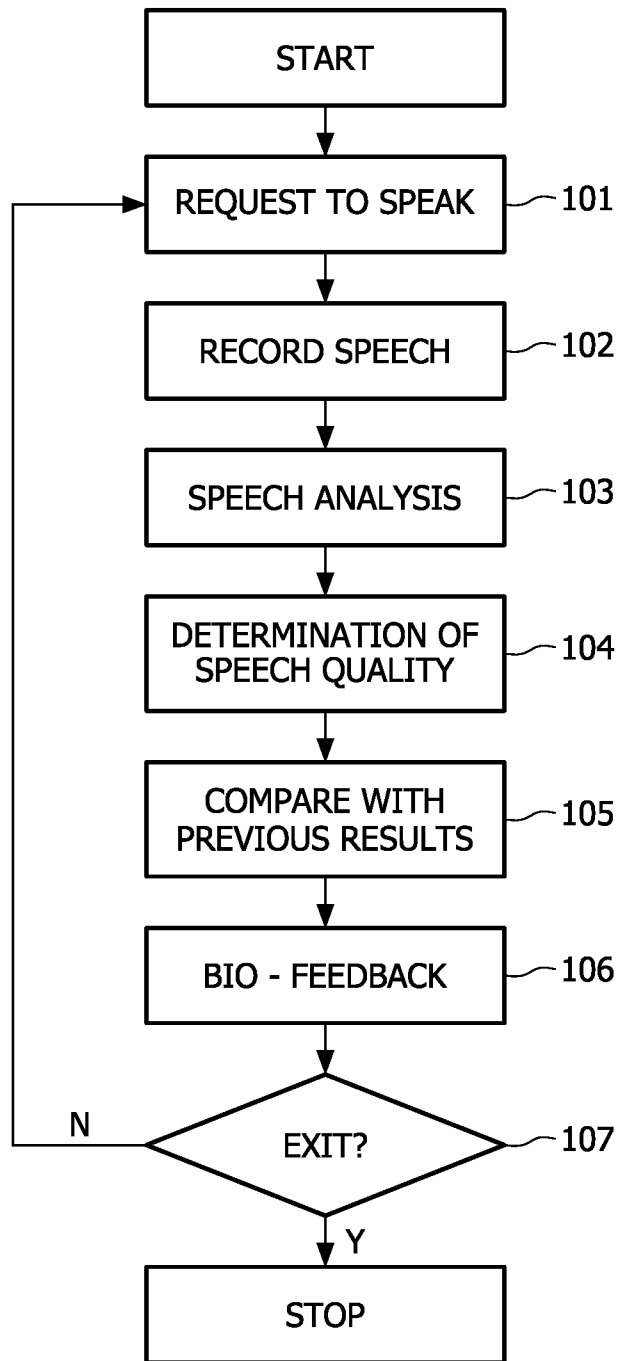15 when the computer program is executed in the computer (2).

1/2

FIG. 1

2/2

START

REQUEST TO SPEAK — 101

RECORD SPEECH — 102

SPEECH ANALYSIS — 103

DETERMINATION OF
SPEECH QUALITY — 104

COMPARE WITH
PREVIOUS RESULTS — 105

BIO - FEEDBACK — 106

N — EXIT? — 107

Y

STOP

FIG. 2

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
INV. G09B19/04

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)

G09B

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

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

EPO-Internal, INSPEC, WPI Data, COMPENDEX, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 1 089 246 A (SIEMENS AG [DE]) 4 April 2001 (2001-04-04) paragraph [0017] - paragraph [0019]	1-8
P,X	US 2007/055523 A1 (YANG GEORGE L [US]) 8 March 2007 (2007-03-08) abstract; figure 5 paragraph [0139]	1,7,8

☐ 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 document 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
8 October 2007	16/10/2007

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Authorized officer Zimmermann, Elko

Form PCT/ISA/210 (second sheet) (April 2005)

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 1089246	A	04-04-2001	DE	19947359 A1	03-05-2001
US 2007055523	A1	08-03-2007	NONE		