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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 November 2007 (01.11.2007)

PCT

(10) International Publication Number
WO 2007/122604 A2

(51) International Patent Classification:
G10L 15/00 (2006.01)

(21) International Application Number:
PCT/IL2007/000179

(22) International Filing Date: 8 February 2007 (08.02.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/411,764 25 April 2006 (25.04.2006) US

(71) Applicant *(for all designated States except US)*: **NICE SYSTEMS LTD.** [IL/IL]; 8 Hapnina Street, 43107 Raanana (IL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **WASSERBLAT, Moshe** [IL/IL]; 14/5 Kikar Malach Hanavi Street, 71700 Modiin (IL). **EILAM, Barak** [IL/IL]; 10 Hapaamonim Street, 43391 Ra'anana (IL).

(74) Agents: **AGMON, Jonathan** et al.; Soroker - Agmon, Advocates & Patent Attorneys, Nolton House, 14 Shenkar Street, 46725 Herzliya Pituach (IL).

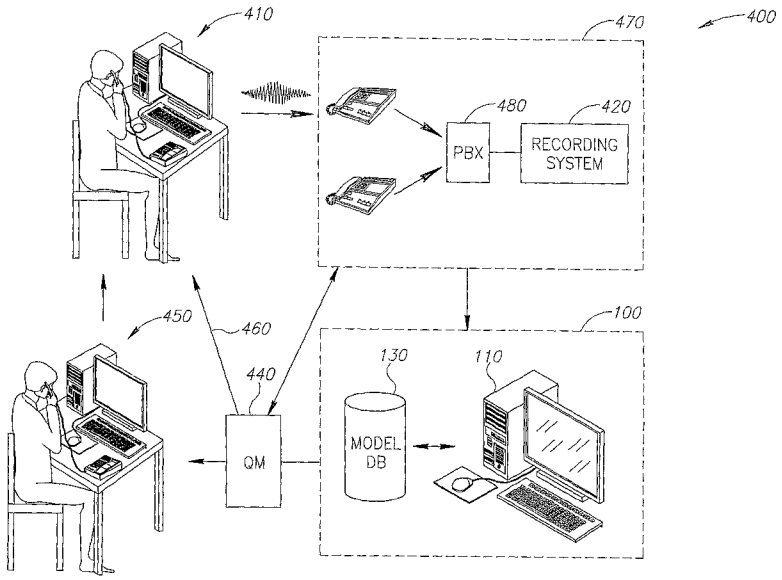
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, 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, 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:
— *without international search report and to be republished upon receipt of that report*

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.

(54) Title: AUTOMATIC SPEECH ANALYSIS



(57) Abstract: A system for providing automatic quality management regarding a level of conformity to a specific accent, including, a recording system, a statistical model database with statistical models representing speech data of different levels of conformity to a specific accent, a speech analysis system, a quality management system. Wherein the recording system is adapted to record one or more samples of a speakers speech and provide it to the speech analysis system for analysis, and wherein the speech analysis system is adapted to provide a score of the speakers speech samples to the quality management system by analyzing the recorded speech samples relative to the statistical models in the statistical model database.

WO 2007/122604 A2

AUTOMATIC SPEECH ANALYSIS

FIELD OF THE INVENTION

The present invention relates generally to automatic association of speech based on speaker characteristics using statistical calculations.

BACKGROUND OF THE INVENTION

Many professions require verbal interaction with clients, for example sales people, bank clerks, and help desk support personnel. Typically, the ability to communicate with the client is not only affected by speaking the same language but also affected by being able to understand the client's accent. In some cases people speaking the same language cannot understand each other because of the accents they are accustomed to. Some accents are considered clearer than others and may be preferred or required for use by people in certain professions, for example television and radio broadcasters.

Nowadays many companies provide telephonic support services, wherein human employees are trained to speak with a clear accent. In some cases these services are outsourced to foreign countries, wherein a foreign language is spoken. The employees performing the service are trained to speak the required language with a desired accent.

Typically a person easily identifies in a short time if another person is speaking with the same accent as accepted in their geographical location or speaks with a different accent. Some people can identify a person's geographical origin based on their accent.

Accent training, and monitoring a trainee's progress is generally expensive and requires individual attention.

US application 10/996,811 filed November 23, 2004, the disclosure of which is incorporated herein by reference, describes a statistical method for speaker spotting in order to split a conversation into separate parts containing the speech of each speaker.

In an article by Yeshwant K. Muthusamy et al. titled "Reviewing Automatic Language Identification" IEEE Signal Processing Magazine October 1994, there is described automated methods of language identification,

In an article by Marc A. Zissman et al. "Comparison of Four Approaches to Automatic Language Identification of Telephone Speech", IEEE Transactions on Speech and Audio Processing, Vol. 4, No. 1, January 1996, there is also described automated methods of language identification.

The above articles describe preparing a statistical model to represent a speech segment or collection of speech segments

In Frederic Bimbot et al., there is described "A Tutorial on Text-Independent Speaker Verification", EURASIP Journal on Applied Signal Processing 2004: 4, 430-451. This article describes the use of statistical method for identifying a user.

Prior art machines dealing with accents are typically used to train a person's accent by requiring the person to repeat a specific word or phrase and comparing the answer to a known digitized pattern. These machines are limited to specific phrases and are not applicable to non pre-selected speech.

SUMMARY OF THE INVENTION

An aspect of an embodiment of the invention relates to a system and method for providing feedback to a speaker regarding their level of conformity to a desired accent. In an exemplary embodiment of the invention, the speaker's speech is recorded and a statistical model is created from the recorded speech. The statistical model is then compared to previously prepared statistical models of speakers with known different levels of conformity to the desired accent, for example wherein the levels are determined by a human expert. Optionally, the system determines the closest statistical model to the recorded speech thus providing a rough level of conformity to the required accent. In an exemplary embodiment of the invention, the system determines a level of closeness of the created statistical model and the determined closest statistical model in order to provide a score regarding the level of closeness of the recorded speech and the desired accent.

Optionally, the score is provided as feedback to the speaker, in order to improve his or her accent. In some embodiments of the invention, the feedback is provided to a supervisor to take remedial actions, for example review the speakers conversations and point out problems to the speaker.

In some embodiments of the invention, the system provides speech segments from the recorded speech that differ the most from the required accent.

A further aspect of an embodiment of the invention, relates to a system and method for automatically identifying a persons accent. In an exemplary embodiment of the invention, a computer receives a digital signal representing a segment of the person's speech. The computer parses the segment and prepares a representation model of the speech segment. The computer compares the representation model to previously stored models representing one or more accents. The computer determines which pre-stored accent model is closest to the model of the received signal and gives a score representing the measure of conformity to it. In some embodiments of the invention, the computer gives

indication regarding segments of the digital signal which do not conform to the determined closest accent.

In an exemplary embodiment of the invention, the automated system queries a caller and analyzes the caller response to determine the accent used by the caller.

In some embodiments of the invention, the automated system is used to train people to speak with a specific accent. In some embodiments of the invention, the automated system gives a score representing the closeness of a monitored person to the desired accent. Optionally, the automated system specifies speech segments, which relatively digress from the determined accent.

In some embodiments of the invention, the automated system operates online in order to give indication in real-time, for example pointing out to a user if he or she digress from the correct accent (e.g. when the user is tired or stressed). Alternatively, the automated system analyzes the received speech segment at a later time and gives indication as a report to the user, an instructor or an employer to determine if the user is progressing over time.

Additionally, an employer can be given immediate feedback or long-term feedback regarding the performance of an employee in conforming to the desired accent.

In some embodiments of the invention, the received segment is pre-processesed by the computer to eliminate effects to the accent determination, for example related to other characteristics of the speaker, for example age group or sex. Alternatively or additionally, the system determines details of the other characteristics (e.g. sex) in addition to the accent.

In an exemplary embodiment of the invention, segments not abiding to the determined accent are automatically replaced or amended to conform to the determined accent.

There is thus provided in accordance with an exemplary embodiment of the invention, a system for providing automatic quality management regarding a level of conformity to a specific accent, comprising:

a recording system;

a statistical model database with statistical models representing speech data of different levels of conformity to a specific accent;

a speech analysis system;

a quality management system;

wherein the recording system is adapted to record one or more samples of a speakers speech and provide it to the speech analysis system for analysis;

wherein the speech analysis system is adapted to provide a score of the speakers speech samples to the quality management system by analyzing the recorded speech samples relative to the statistical models in the statistical model database. Optionally, analysis of the recorded speech samples by the speech analysis system comprises:

preparing a statistical model from the speech samples; and

comparing it to other statistical models in the statistical model database to determine the closest model and a level of conformity to the closest model.

In an exemplary embodiment of the invention, the quality management system is adapted to provide feedback to the speaker. Optionally, the quality management system is adapted to provide feedback to a supervisor. In an exemplary embodiment of the invention, the quality management system is adapted to provide feedback regarding the performance of the speaker based on the provided score and previous provided scores of the speaker. Optionally, the feedback further comprises specific speech segments with a maximum level of deviation from the model which most conforms to a specific accent.

There is thus additionally provided in accordance with an exemplary embodiment of the invention, a method of providing automated feedback to a speaker regarding conformity of their accent to a specific accent, comprising:

recording one or more samples of the speakers speech;

producing a statistical model from said speech samples;

comparing the produced statistical model to previously prepared statistical models of speech samples with different levels of conformity to a specific accent;

determining the model that conforms the best to the produced statistical model and a level of conformity to the best conforming model; and

providing as feedback a score for the speech samples regarding its conformity to the specific accent based on the determining. Optionally, the feedback further comprises specific speech segments with a maximum level of deviation from the best conforming model. In an exemplary embodiment of the invention, the feedback takes into account scores of the speaker from previous recordings.

There is thus additionally provided in accordance with an exemplary embodiment of the invention, a method of automatic accent identification for quality management comprising:

creating one or more statistical accent models representing accents from one or more collections of training speech data;

inputting a speech signal for analysis;

preparing a statistical speech model representing the input speech signal;

comparing the statistical speech model with the one or more statistical accent models;

calculating a score resulting from the comparison of the statistical speech model with each statistical accent model;

determining a closest statistical accent model to the statistical speech model; and

providing the scores to a quality management system to provide feedback.

Optionally, the determining is performed substantially in real-time. In some embodiments of the invention, the method further comprises giving indication regarding segments of the speech signal which do not conform to the determined closest statistical accent module. Optionally, the indication is given substantially in real-time. In an exemplary embodiment of the invention, the quality management system provides indication regarding the quality of the accent of a user. Optionally, the method further comprises notifying a user if the determined closest statistical accent changes during a conversation. In an exemplary

embodiment of the invention, the determining further determines other characteristics of a speaker. Optionally, the one or more statistical models are updated based on speech signals from groups of substantially equal scoring users.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood and appreciated more fully from the following detailed description taken in conjunction with the attached drawings. Identical structures, elements or parts, which appear in more than one figure, are generally labeled with a same or similar number in all the figures in which they appear, wherein:

Fig. 1 is a schematic illustration of a system for analyzing speech to determine the accent used by the speaker, according to an exemplary embodiment of the invention;

Fig. 2 is a flow diagram of a method of determining an accent using an automated system, according to an exemplary embodiment of the invention;

Fig. 3 is a schematic illustration of analysis of a speech signal by the system for analyzing speech, according to an exemplary embodiment of the invention; and

Fig. 4 is a schematic illustration of a system for analyzing a speaker's accent and providing feedback, according to an exemplary embodiment of the invention.

DETAILED DESCRIPTION

Fig. 1 is a schematic illustration of a system 100 for analyzing speech to determine the accent used by the speaker, according to an exemplary embodiment of the invention. In an exemplary embodiment of the invention, system 100 comprises tools for performing speech analysis, for example using a computer 110 and a database 130. In an exemplary embodiment of the invention, system 100 is trained to recognize one or more different accents to which it can compare an input speech signal 170. Optionally, one or more collections of speech data 120, each containing speech samples of a specific accent, are provided to computer 110 in order to build statistical models 160 for comparing with input speech signal 170.

In some embodiments of the invention, multiple collections of speech data 120 are used to create multiple models for each accent, representing people with different speech characteristics, for example men and women, children and grown ups. Alternatively the models are created by extracting accent related details from collections of speech data 120 and the models are normalized relative to other differences, for example level of conformity to a specific accent.

In an exemplary embodiment of the invention, computer 110 using an analysis program accepts a collection of speech data 120 as a digital signal 140 and parses it into small segments 180, for example each of duration of about 10ms. Optionally, a virtual window 150 of a size larger than segments 180 (for example twice as big), is used to inspect each segment. Optionally, window 150 detects the basic accent related details in digital signal 140 and creates a statistical model 160 which represents the frequency of each type of sound or statistical likelihood of occurrence of sound combinations in the accent being inspected.

In an exemplary embodiment of the invention, model 160 is created based on details such as:

Phonology – phone sets used by the accent.

Acoustic phonetics – frequency of occurrence of phones.

Phonotactics – rules governing combinations of phones.

Prosody – duration, pitch, and stress of the sounds used and their frequency of appearance.

Optionally, statistical models 160 are stored in database 130, internal to computer 110. Alternatively, database 130 is external to computer 110, for example on a network server.

In some embodiments of the invention, the building process of statistical models 160 is performed using learning algorithms, for example neural networks or Hidden Markov Models (HMM) similar to the methods used for language determination as described in the articles mentioned above.

Fig. 2 is a flow diagram of a method 200 of determining an accent using automated system 100, according to an exemplary embodiment of the invention, and Fig. 3 is a schematic illustration of the process of analysis of an input speech signal 170 by system 100, according to an exemplary embodiment of the invention.

In an exemplary embodiment of the invention, after training system 100 and creating a set of reference statistical models 160, system 100 is ready to analyze input speech signal 170 and determine if it conforms to any of models 160. Optionally, system 100 inputs (210) input speech signal 170 for analysis. System 100 processes input speech signal 170 and prepares (220) a model representing input speech signal 170 according to a similar process as described above for preparing the reference statistical models 160. Optionally, system 100 compares (230) the model created for input speech signal 170 with models 160 and produces (240) a score 320 representing a measure of conformity of input speech signal 170 to models 160.

In an exemplary embodiment of the invention, system 100 determines (250) the accent whose model 160 results in a score 320 with the closest match 330 to the model of input speech signal 170. In an exemplary embodiment of the invention, the comparison and selection is performed using Gaussian Mixture Models (GMM) as is known in the art. Alternatively, other statistical methods can be used to determine the closest matching model 330 to input speech signal 170.

Optionally, a threshold value is selected, so that if input speech signal 170 is not close enough to any of the models, no model will be selected as representing the closest accent.

In some embodiments of the invention, once determining the closest accent, to input speech signal 170, system 100 determines segments of input speech signal 170, which do not conform to the accent, for example due to errors or inability of the speaker. In some embodiments of the invention, system 100 reports these errors for future training of the speaker. Alternatively, in a real-time system, system 100 may amend known errors by replacing specific waveforms of the digital representation with an amended waveform.

In an exemplary embodiment of the invention, during a conversation, system 100 analyzes the speech of a single participant. Alternatively, system 100 analyzes the entire conversation and splits the speech signal to a segment for each speaker according to methods for example as described in the application incorporated above by reference.

In some embodiments of the invention, system 100 analyzes speech data in real time, and continuously determines the closest accent relative to input speech signal 170. Optionally, the determined closest accent may change responsive to changes in the accent used by the speaker being monitored.

In some embodiments of the invention, every sentence or a speech segment of a preset time interval is analyzed to determine the closest accent. Alternatively, the beginning of a conversation is analyzed to determine the closest accent and the continuation is analyzed for conformity to the determined accent.

In an exemplary embodiment of the invention, system 100 serves to identify a caller's accent and select a call responder that matches the callers accent. Alternatively or additionally, system 100 serves to monitor a call responder and give indication if the responder is digressing from the required accent.

In some embodiments of the invention, system 100 queries the caller with questions that will invoke answers that will assist the system in determining

the caller's accent, for example by asking the caller a question that requires a long response. In some embodiments of the invention, the caller can be asked to repeat a specific sentence with pre selected sounds that are known to differentiate between different accents of a specific language in order to shorten the determination process and/or assure its success.

In some embodiments of the invention, system 100 is used to train a speaker to speak with a specific accent. In some embodiments of the invention, system 100 is used to test a speaker regarding his or her ability to speak with a certain accent, for example to screen job applicants or rank employees according to speech clarity.

In some embodiments of the invention, system 100 is used to track progress of a trainee in learning a new accent or improving their accent.

Fig. 4 is a schematic illustration of a system 400 for analyzing a speaker's accent and providing feedback, according to an exemplary embodiment of the invention. In an exemplary embodiment of the invention, an agent 410 accepts calls from people and provides verbal assistance, for example wherein agent 410 serves as a customer service representative or a help line technician. Optionally, the conversations of the agent and other agents are controlled by a call contact center 470, which controls the reception of calls by agents and records the calls. In an exemplary embodiment of the invention, call contact center includes a private branch exchange (PBX) controller 480 to navigate calls from customers to the receiving agents. Additionally, call contact center 470 includes a recording system 420 to record the conversations between the clients and the agents.

In an exemplary embodiment of the invention, system 100 is used to analyze the speech data as described above. Optionally, computer 110 of system 100 is used to analyze the speech data with the aid of model database 130 that stores predetermined accent models. In an exemplary embodiment of the invention, computer 110 receives recorded conversations from recording system 420 to analyze an agent's speech for conformity to the predetermined models of the required accent and other accents.

In an exemplary embodiment of the invention, multiple models are prepared for a specific accent and stored in model database 110 of system 100 as described above. Each of the multiple models is based on a pre-graded selection of speech data. Optionally, a supervisor gives a score for many speech sessions of agents, for example the score values can be 100%, 90%, 80% and so forth. Computer 110 prepares a model for each score (100%, 90% ...) and these models are used in the accent determination by computer 110. In an exemplary embodiment of the invention, computer 110 determines a score based on the score of the model which is closest to the speech sample being evaluated. Additionally, computer 110 determines a level of conformity to the closest model. Optionally, these values are used to provide a score for the evaluated speech sample. In some embodiments of the invention, computer 110 additionally, compares the speech sample with the best model and provides exemplary speech segments from the evaluated speech sample, which deviate the most from the best model.

In an exemplary embodiment of the invention, the determined scores and samples are provided to a quality management system 440 (e.g. comprising a computer and a database) and are accumulated for each agent 410 over a period of time, for example a few days or a few weeks. Optionally, quality management system 440 evaluates an agents results over a period of time, to determine the agents weak points and if the agent is improving or not. Optionally, quality management system 440 provides the information as feedback 460 to a supervisor 450 and/or to agent 410 so that agent 410 and/or supervisor 450 can take remedial actions, for example providing additional training regarding the agents weak points.

In some embodiments of the invention, the computer evaluation can be performed online during the duration of the conversation conducted by agent 410 and give immediate feedback taking into account previous evaluations, for example computer 110 can provide a visual feedback indicating a level of deviation of the agent from the required accent. Optionally, quality management

system 440 may be connected directly to call contact center 470 in order to extract specific speech segments for providing to supervisor 450 or agent 410.

It should be appreciated that the above described methods and apparatus may be varied in many ways, including omitting or adding steps, changing the order of steps and the type of devices used. It should be appreciated that different features may be combined in different ways. In particular, not all the features shown above in a particular embodiment are necessary in every embodiment of the invention. Further combinations of the above features are also considered to be within the scope of some embodiments of the invention.

Section headings are provided for assistance in navigation and should not be considered as necessarily limiting the contents of the section.

It will be appreciated by persons skilled in the art that the present invention is not limited to what has been particularly shown and described hereinabove. Rather the scope of the present invention is defined only by the claims, which follow.

CLAIMS

1. A system for providing automatic quality management regarding a level of conformity to a specific accent, comprising:

a recording system;

a statistical model database with statistical models representing speech data of different levels of conformity to a specific accent;

a speech analysis system;

a quality management system;

wherein said recording system is adapted to record one or more samples of a speakers speech and provide it to said speech analysis system for analysis;

wherein said speech analysis system is adapted to provide a score of the speakers speech samples to said quality management system by analyzing said recorded speech samples relative to the statistical models in said statistical model database.

2. A system according to claim 1, wherein analysis of said recorded speech samples by said speech analysis system comprises:

preparing a statistical model from said speech samples; and

comparing it to other statistical models in said statistical model database to determine the closest model and a level of conformity to the closest model.

3. A system according to claim 1, wherein said quality management system is adapted to provide feedback to said speaker.

4. A system according to claim 1, wherein said quality management system is adapted to provide feedback to a supervisor.

5. A system according to claim 1, wherein said quality management system is adapted to provide feedback regarding the performance of the speaker based on the provided score and previous provided scores of the speaker.

6. A system according to claim 1, wherein said feedback further comprises specific speech segments with a maximum level of deviation from the model which most conforms to a specific accent.

7. A method of providing automated feedback to a speaker regarding conformity of their accent to a specific accent, comprising:

recording one or more samples of the speakers speech;

producing a statistical model from said speech samples;

comparing said produced statistical model to previously prepared statistical models of speech samples with different levels of conformity to a specific accent;

determining the model that conforms the best to said produced statistical model and a level of conformity to the best conforming model; and

providing as feedback a score for said speech samples regarding its conformity to the specific accent based on said determining.

8. A method according to claim 7, wherein said feedback further comprises specific speech segments with a maximum level of deviation from the best conforming model.

9. A method according to claim 7, wherein said feedback takes into account scores of the speaker from previous recordings.

10. A method of automatic accent identification for quality management comprising:

creating one or more statistical accent models representing accents from one or more collections of training speech data;

inputting a speech signal for analysis;

preparing a statistical speech model representing the input speech signal;

comparing the statistical speech model with said one or more statistical accent models;

calculating a score resulting from the comparison of said statistical speech model with each statistical accent model;

determining a closest statistical accent model to said statistical speech model; and

providing the scores to a quality management system to provide feedback.

11. A method according to claim 10, wherein said determining is performed substantially in real-time.

12. A method according to claim 10, further comprising giving indication regarding segments of said speech signal which do not conform to the determined closest statistical accent module.

13. A method according to claim 12, wherein said indication is given substantially in real-time.

14. A method according to claim 10, wherein said quality management system provides indication regarding the quality of the accent of a user.

15. A method according to claim 10, further comprising notifying a user if the determined closest statistical accent changes during a conversation.

16. A method according to claim 10, wherein said determining further determines other characteristics of a speaker.

17. A method according to claim 10, wherein said one or more statistical models are updated based on speech signals from groups of substantially equal scoring users.

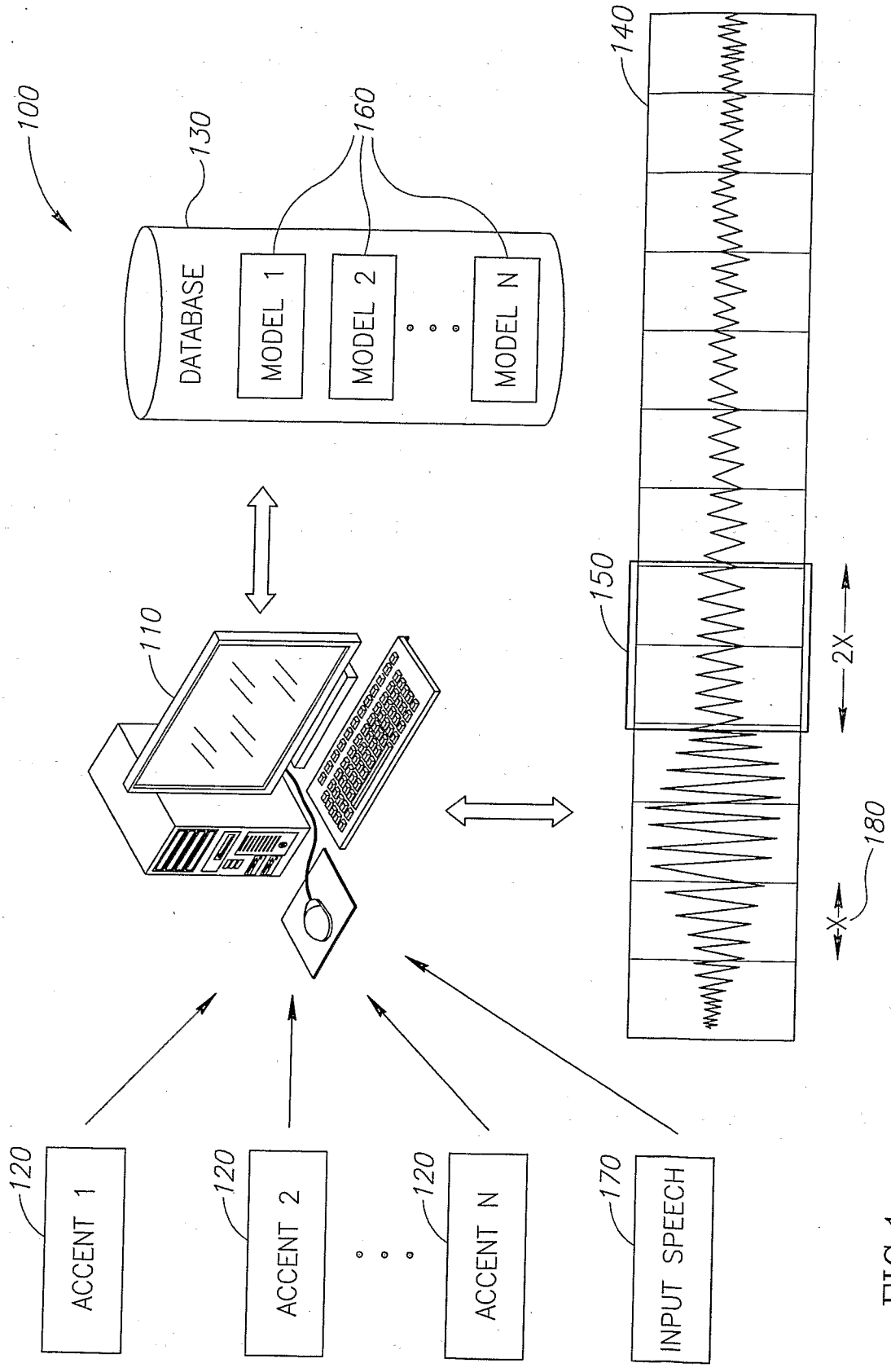

100
130
160
140
DATABASE
MODEL 1
MODEL 2
MODEL N
110
150
2X
180
X
120
ACCENT 1
120
ACCENT 2
120
ACCENT N
170
INPUT SPEECH

FIG.1

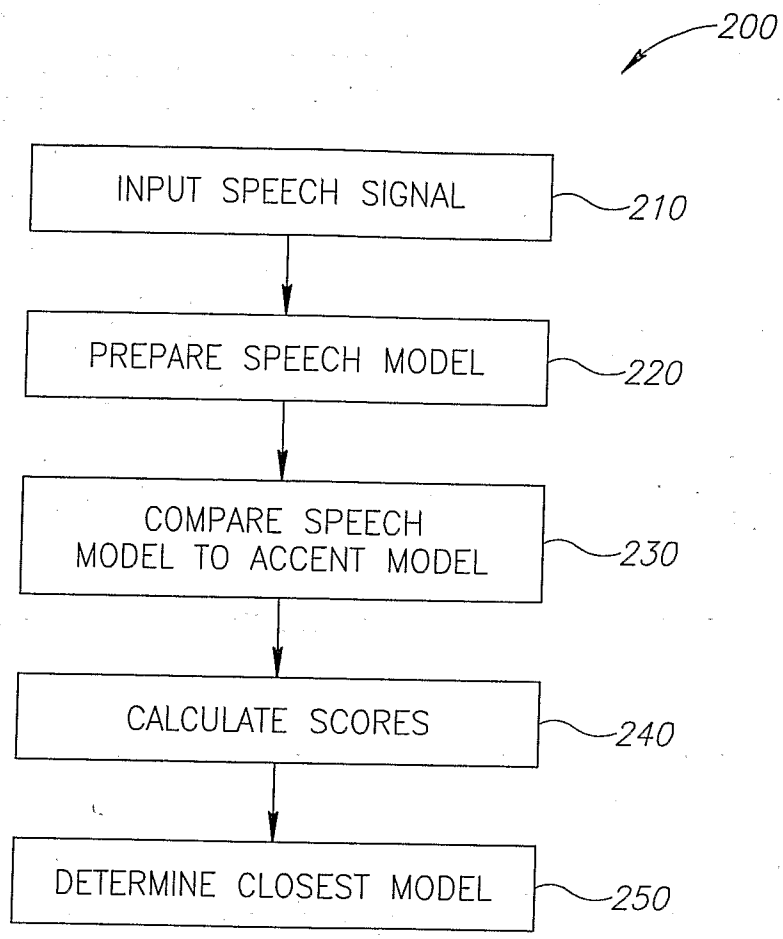

200
INPUT SPEECH SIGNAL
210
PREPARE SPEECH MODEL
220
COMPARE SPEECH MODEL TO ACCENT MODEL
230
CALCULATE SCORES
240
DETERMINE CLOSEST MODEL
250

FIG.2

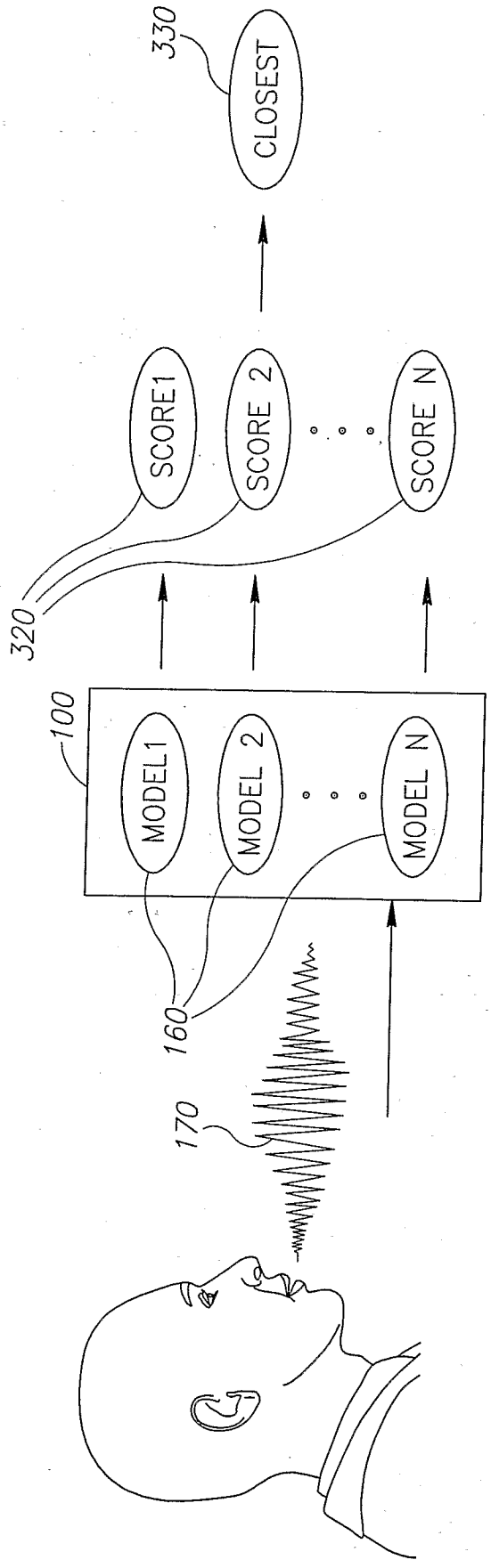

330
CLOSEST
SCORE1
SCORE 2
SCORE N
320
100
MODEL1
MODEL 2
MODEL N
160
170

FIG.3

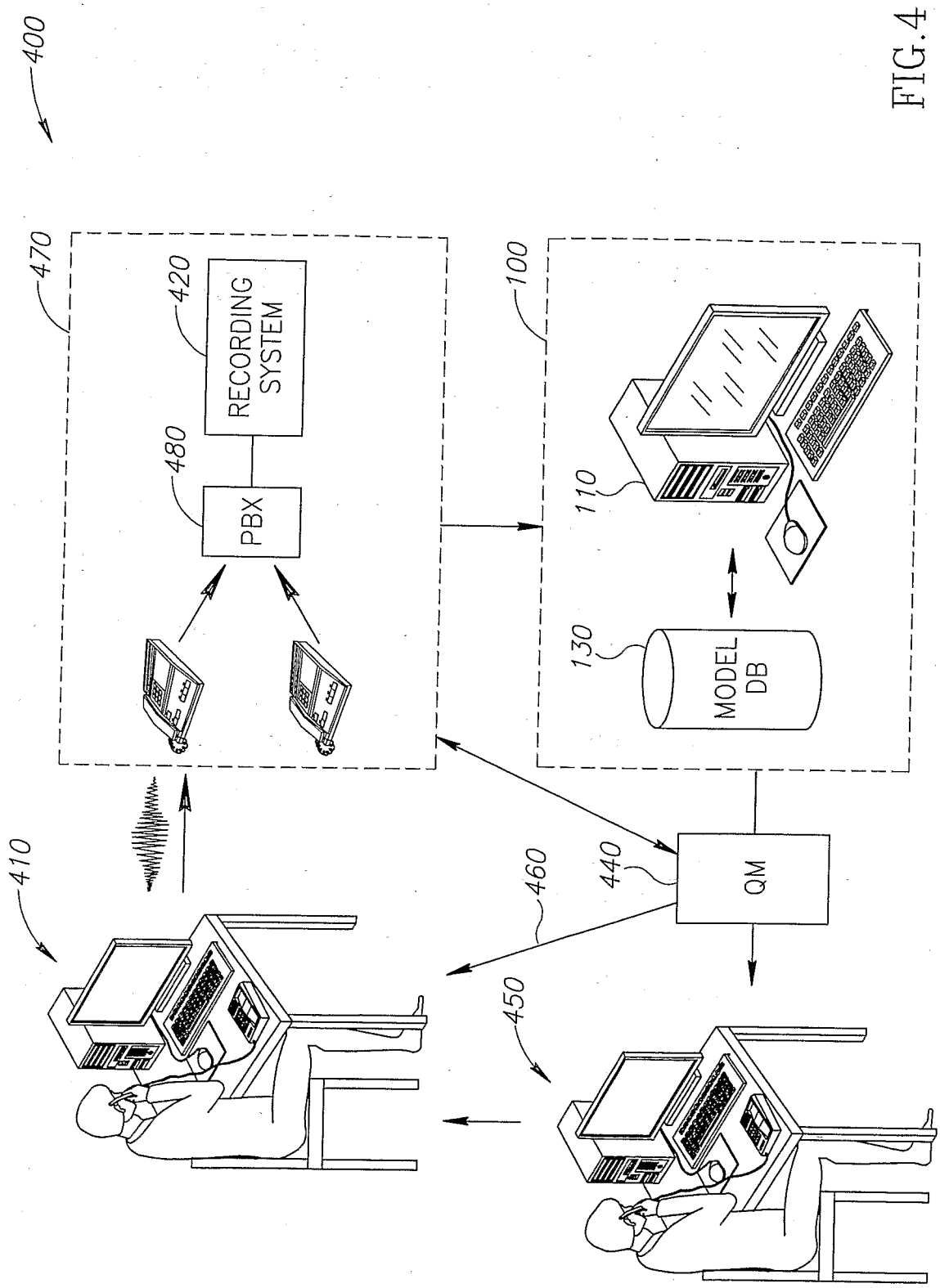

400
470
420
RECORDING SYSTEM
480
PBX
100
110
130
MODEL DB
410
460
440
QM
450

FIG.4