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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 July 2006 (20.07.2006)

PCT

(10) International Publication Number
WO 2006/076280 A2

(51) International Patent Classification:
G10L 17/00 (2006.01)

(21) International Application Number:
PCT/US2006/000646

(22) International Filing Date: 10 January 2006 (10.01.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/643,131 11 January 2005 (11.01.2005) US

(71) Applicant (*for all designated States except US*): EDUCATIONAL TESTING SERVICE [US/US]; Rosedale Road, Princeton, NJ 08541 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): HIGGINS, Derrick [US/US]; 111 N. Second Avenue, Highland Park, NJ 08904 (US). ZECHNER, Klaus [US/US]; 10 Pelham Street, Prnceton, NJ 08540 (US). FUTAGI, Yoko [JP/US]; 3 Karena Lane, Lawrenceville, NJ 08648 (US). LAWLESS, Rene [US/US]; 50 Titus Mill Road, Pennington, NJ 08534 (US).

(74) Agent: MELNIK, W., Joseph; Pepper Hamilton LLP, One Mellon Center, 50th Floor, 500 Grant Street, Pittsburgh, PA 15219 (US).

(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, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

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:** METHOD AND SYSTEM FOR ASSESSING PRONUNCIATION DIFFICULTIES OF NON-NATIVE SPEAKERS

(57) **Abstract:** The present disclosure presents a useful metric for assessing the relative difficulty which non-native speakers face in pronouncing a given utterance and a method and systems for using such a metric in the evaluation and assessment of the utterances of non-native speakers. In an embodiment, the metric may be based on both known sources of difficulty for language learners and a corpus-based measure of cross-language sound differences. The method may be applied to speakers who primarily speak a first language speaking utterances in any non-native second language.

WO 2006/076280 A2

METHOD AND SYSTEM FOR ASSESSING PRONUNCIATION DIFFICULTIES OF NON-NATIVE SPEAKERS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This patent application claims priority to, and incorporates by reference in its entirety, U.S. Provisional Patent Application No. 60/643,131, entitled "Method and System for Assessing Pronunciation Difficulties of Non-Native Speakers" and filed January 11, 2005.

TECHNICAL FIELD

[0002] The present invention relates generally to the field of speech recognition and assessment. The present invention particularly relates to a method and system for quantifying the relative pronunciation difficulty that an utterance in a second language poses for a non-native speaker.

BACKGROUND

[0003] Much literature exists on the specific aspects of native language sound structure that are difficult for non-native speakers, and to what extent features of a native language may interfere with learning correct pronunciation in a second language. This work is rooted in Uriel Weinreich's notion of interference and is developed in more recent works into Terence Odlin's and Larry Selinker's concepts of language transfer and interlanguage. "*Languages in Contact*," *Uriel Weinreich*, (1968); "*Language Transfer*," *Terence Odlin*, (1989); "*Rediscovering Interlanguage*," *Larry Selinker*, (1992). From the pedagogical side, Avery and Ehrlich's "*Teaching America English Pronunciation*" and Celce – Murcia et al's "*Teaching Pronunciation*" are two reference materials which provide insight into the structure of English phonology and phonetic's, the implications it has for the learning of pronunciation, and how it interacts with a speakers' linguistic backgrounds. "*Teaching American English Pronunciation*," *Peter Avery and Susan Ehrlich*, (1992); "*Teaching Pronunciation,*" *Marianne Celce – Murcia, Donna M. Brinton, and Janet M. Goodwin*, (1996). However, none of the literature quantifies the difficulty that a non-native speaker of a particular non-native language background would have in pronouncing a given utterance of the native language.

[0004] The present invention is directed to solving one or more of the above-listed problems.

SUMMARY

[0005] In an embodiment, a method of assessing the pronunciation difficulties of a non-native speaker may include determining one or more sources of the pronunciation difficulties between a language of the non-native speaker and a second language, assigning a weight to each source, calculating a phonetic difficulty score based on the one or more sources and the weight assigned to each source, calculating a language model score based on a sound comparison between the language of the non-native speaker and the second language, normalizing the phonetic difficulty score and the language model score, and calculating a metric by calculating the sum of the normalized phonetic difficulty score and the normalized language model score.

[0006] In an embodiment, calculating the phonetic difficulty score may include calculating a product of the number of times the source occurs in the language of the non-native speaker and the weight of the source. Calculating the language model score may include calculating a language model for the language of the non-native speaker and the second language and calculating a cross-entropy of an utterance with respect to the language model of the non-native speaker's first language inversely weighted by the cross-entropy of the utterance with respect to the language model of the second language. A language model may be calculated by phonetically transcribing the language of the non-native speaker and the second language. The cross-entropy of an utterance with respect to the language model of the non-native speaker may be calculated by assigning a lower score to utterances of the second language that are similar to sounds of the first language and assigning a higher score to utterances of the second language that are not similar to sounds of the first language. A normalization equation may be used to normalize the phonetic difficulty score and the language model score. The normalized phonetic difficulty score and the normalized language model score may have the same mean and standard deviation once each has been normalized.

[0007] In an embodiment, a system may include a processor, a processor-readable storage medium in communication with the processor, and a display. The system may be used to assess the pronunciation difficulties of a non-native speaker. The processor-readable storage medium may contain one or more programming instructions for performing the method of assessing the pronunciation difficulties of a non-native speaker.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Aspects, features, benefits, and advantages of the embodiments of the present invention will be apparent with regard to the following description, appended claims, and accompanying drawings where:

[0009] FIG. 1 depicts an exemplary method of assessing pronunciation difficulties for non-native speakers according to an embodiment.

[0010] FIG. 2 lists exemplary sources of difficulty that a native speaker of Japanese may have when speaking English.

[0011] FIG. 3 depicts an exemplary system for assessing pronunciation difficulties according to an embodiment.

DETAILED DESCRIPTION

[0012] Before the present methods, systems, and materials are described, it is to be understood that this invention is not limited to the particular methodologies, systems, and materials described, as these may vary. It is also to be understood that the terminology used in the description is for the purpose of describing the particular versions or embodiments only, and is not intended to limit the scope of the invention, which will be limited only by the appended claims.

[0013] It must also be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural references unless the context clearly dictates otherwise. Thus, for example, reference to a "source" is a reference to one or more sources and equivalents thereof known to those skilled in the art, and so forth. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art. Although any methods, materials, and devices similar or equivalent to those described herein can be used in the practice or testing of embodiments of the invention, the preferred methods, materials, and devices are now described. All publications mentioned herein are incorporated by reference. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

[0014] FIG. 1 shows an exemplary method of assessing pronunciation difficulties for non-native speakers. A determination of one or more sources of the pronunciation difficulties between a language of the non-native speaker and a second language, as represented by **10**, may be made by reviewing studies from language pedagogy, phonetics, and interlanguage phonology, or may be based on an observation of learner tendencies. These sources of difficulty may include, without limitation, shibboleths, which are well known markers of non-native speakers with specific backgrounds. For example, the well-known tendency of Japanese speakers to have difficulty with the contrastive "r" and "l" sounds in the English language is reported in many studies. Other, less obvious tendencies may be found in the literature, such as the observation that Japanese speakers may have difficulty with English voiceless stops, producing them with a shorter voice-onset time than native speakers of the English language.

[0015] Once a determination of one or more of the sources has been made, each source may then be assigned a weight **12** that may be a measure of the severity of the source as reflected in the pedagogical and linguistic literature related to the speaker's background. In addition, the weights may be chosen in consultation with a native speaker of the first language. FIG. 2 lists exemplary sources of difficulty that a native speaker of Japanese may have with the English language. A description of the source is provided along with a symbol that may be used to represent the source. Each source also includes a weight.

[0016] After determining the one or more sources and assigning a weight to each source, a phonetic difficulty score may be calculated based on the one or more sources and the weight assigned to each source **14**. The formula for the phonetic difficulty score may comprise calculating a product of the number of times the source occurs in an utterance of the non-native speaker and the weight of the source. More explicitly, the phonetic difficulty score may be defined by the following equation:

$$Score_{phonetic difficulty} = \sum_{\forall t} Count(t,u) \cdot W(t)$$

where: t is the phonetic source,

u is the utterance,

$W(t)$ is the weight assigned to the source, and

$Count(t,u)$ is the number of times t appears in u.

[0017] The phonetic difficulty score may be tied to actual error types that language learners of a given first language background commit in speaking the second language. Since the existence of a difference between languages does not guarantee interference effects, empirical data may be used to determine whether a particular error type occurs with some frequency. However, the phonetic difficulty score alone may be limited because it may require the advance detection of error types. Moreover, identified errors alone may not determine the relative importance of each source of difficulty.

[0018] After calculating the phonetic difficulty score, a language model score may then be calculated based on a sound comparison between the language of the non-native speaker and a second language **16**. In an embodiment, calculating the language model score may comprise calculating a language model for the language of the non-native speaker and the second language and calculating a cross-entropy of an utterance with respect to the language model of the non-native speaker's first language inversely weighted by the cross-entropy of the utterance with respect to the language model of the second language. The language model score may be derived from the following equation:

$$D(u, P_{L1}[\cdot] \,\|\, P_{L2}[\cdot]) = H(u, P_{L1}[\cdot]) - H(u, P_{L2}[\cdot])$$

$$= -\log\left(\frac{P_{L1}[u]}{P_{L2}[u]}\right)$$

Where: u is the utterance,

$P_{L1}[\cdot]$ is the distribution defined by the language model of the first language,

$P_{L2}[\cdot]$ is the distribution defined by the language model of the second language,

$H(u, P[\cdot])$ represents the cross entropy of the utterance u with respect to $P[\cdot]$, and

D indicates that the language model score is a form of the KL divergence.

[0019] The language model for the language of the non-native speaker and the second language may comprise phonetically transcribing the language of the non-native speaker and the second language. These languages may be phonetically transcribed by applying a phonetic dictionary to the text of a collection of recorded utterances of each language. The cross-entropy of an utterance with respect to the language model of the non-native speaker's first language may be calculated by assigning a lower score to utterances of the second language that are similar to sounds of the first language and assigning a higher score to utterances of the second language that are not similar to sounds of the first language. The cross-entropy may be determined by the following equation:

$$H(u, P_{L1}[\cdot]) = -\log(P_{L1}[u])$$

[0020] Similarly, the cross entropy of the utterance with respect to the second language would be calculated as:

$$H(u, P_{L2}[\cdot]) = -\log(P_{L2}[u])$$

[0021] Unlike the phonetic difficulty score, the language model may use observed differences between languages and not observed sources of difficulty. The language model score may be based on statistical models of the sound structure of the first and second languages. Given these models, a metric that gives high values to sound patterns that are uncommon in the first language, but relatively common in the second language, may be designed.

[0022] After calculating both of the scores, the scores may be normalized **18**, so that they make approximately the same contribution to a final difficulty score. The normalized scores may comprise the same mean and standard deviation. Each score may be normalized by using the following normalization equation:

$$\mathbf{Norm}(x) =_{\text{def}} \frac{x - \overline{X}}{\sigma_x}$$

where: X represents the distribution of scores

$\overline{X}$ represents the mean of this distribution

σ_X represents the standard deviation of this distribution, and

x is a score drawn from this distribution.

[0023] Once the scores have been normalized, a metric may be calculated by adding the normalized phonetic difficulty score and a normalized language difficulty score **20**. The metric may quantify the pronunciation difficulty that an utterance in a second language poses for non-native speakers of different first language backgrounds. The metric may combine information from a measure of cross-language sound differences and the literature of language pedagogy to produce a single prediction of pronunciation difficulty. The method may be applicable to any pair of languages that could serve as the first and second languages.

[0024] FIG. 3 shows a block diagram of exemplary hardware that may be used to contain and/or implement the program instructions of system embodiments of the present invention. A bus 228 may serve as the information highway interconnecting the other illustrated components of the hardware. A processor **202**, such as a central processing unit

(CPU), may perform calculations and logic operations required to execute a program. A processor-readable storage medium, such as read only memory (ROM) **218** and/or random access memory (RAM) **220**, may be in communication with the processor **202** and may contain one or more programming instructions for performing the method of assessing the pronunciation difficulties of a non-native speaker. Optionally, program instructions may be stored on a computer readable carrier such as a digital disk, recordable memory device, or other recording medium, a communications signal, or a carrier wave.

[0025] A disk controller **204** interfaces one or more optional disk drives to the system bus **228**. These disk drives may be external or internal floppy disk drives such as **210**, external or internal CD-ROM, CD-R, CD-RW or DVD drives such as **206**, or external or internal hard drives **208**. As indicated previously, these various disk drives and disk controllers are optional devices

[0026] Each of the element managers, real-time data buffers, conveyors, file input processor, database index shared access memory loader, reference data buffer and data managers may include a software application stored in one or more of the disk drives connected to the disk controller **204**, the ROM **218** and/or the RAM **220**. Preferably, the processor **202** may access each component as required.

[0027] A display interface **222** may permit information from the bus **228** to be displayed on a display **224** in audio, graphic, or alphanumeric format. Communication with external devices may optionally occur using various communication ports **226**.

[0028] In addition to the standard computer-type components, the hardware may also include data input devices, such as a keyboard **214**, or other input device **216**, such as a microphone, remote control, pointer, mouse and/or joystick.

[0029] It is to be understood that the invention is not limited in its application to the details of construction and to the arrangements of the components or steps set forth in this description or illustrated in the drawings. The disclosed method and system are capable of other embodiments and of being practiced and carried out in various ways. Hence, it is to be understood that the phraseology and terminology employed herein are for the purpose of description and should not be regarded as limiting.

[0030] As such, those skilled in the art will appreciate that the conception upon which this disclosure is based may readily be utilized as a basis for the designing of other structures, methods and systems for carrying out the several purposes of the present invention. It is important, therefore, that the claims be regarded as including such equivalent constructions insofar as they do not depart from the spirit and scope of the disclosed embodiments.

We claim:

1. A method of assessing pronunciation difficulties of a non-native speaker, the method comprising:

determining one or more sources of pronunciation difficulties between a language of a non-native speaker and a second language;

assigning a weight to each source;

calculating a phonetic difficulty score based on the one or more sources and the weight assigned to each source;

calculating a language model score based on a sound comparison between the language of the non-native speaker and the second language;

normalizing the phonetic difficulty score and the language model score; and

calculating a metric from the normalized phonetic difficulty score and the normalized language model score.

2. The method of claim 1, wherein calculating the phonetic difficulty score comprises calculating a product of the number of times the source occurs in an utterance of the non-native speaker and the weight of the source.

3. The method of claim 1, wherein calculating the language model score comprises:

calculating a language model for the language of the non-native speaker and the second language; and

calculating a cross-entropy of an utterance with respect to the language model of the language inversely weighted by the cross-entropy of the utterance with respect to the language model of the second language.

4. The method of claim 3, wherein calculating a language model for the language of the non-native speaker and the second language comprises phonetically transcribing the first language of the non-native speaker and the second language.

5. The method of claim 3, wherein calculating the cross-entropy of an utterance with respect to the language model of the first language comprises:

assigning a lower score to utterances of the second language that are similar to sounds of the language of the non-native speaker; and

assigning a higher score to utterances of the second language that are not similar to sounds of the language of the non-native speaker.

6. The method of claim 1, wherein normalizing the phonetic difficulty score and the language model score comprises determining a normalized phonetic difficulty score and a normalized language model score such that the mean of the normalized phonetic difficulty score is equal to the mean of the normalized language model score and the standard deviation of the normalized phonetic difficulty score is equal to the standard deviation of the normalized language model score.

7. The method of claim 1, wherein calculating the metric comprises calculating the sum of the normalized phonetic difficulty score and the normalized language model score.

8. A system for assessing the pronunciation difficulties of a non-native speaker comprising:

a processor; and

a processor-readable storage medium in communication with the processor,

wherein the processor-readable storage medium contains one or more programming instructions for performing a method of assessing the pronunciation difficulties of a non-native speaker, the method comprising:

determining one or more sources of pronunciation difficulties between a language of a non-native speaker and a second language,

assigning a weight to each source,

calculating a phonetic difficulty score based on the one or more sources and the weight assigned to each source,

calculating a language model score based on a sound comparison between the language of the non-native speaker and the second language,

normalizing the phonetic difficulty score and the language model score, and

calculating a metric from the normalized phonetic difficulty score and the normalized language model score.

9. The system of claim 8, wherein calculating the phonetic difficulty score comprises calculating a product of the number of times the source occurs in an utterance of the non-native speaker and the weight of the source.

10. The system of claim 8, wherein calculating the language model score comprises:

calculating a language model for the language of the non-native speaker and the second language; and

calculating a cross-entropy of an utterance with respect to the language model of the language of the non-native speaker inversely weighted by the cross-entropy of the utterance with respect to the language model of the second language.

11. The system of claim 10, wherein calculating a language model for the language of the non-native speaker and the second language comprises phonetically transcribing the language of the non-native speaker and the second language.

12. The system of claim 10, wherein calculating the cross-entropy of an utterance with respect to the language model of the first language comprises:

assigning a lower score to utterances of the second language that are similar to sounds of the language of the non-native speaker; and

assigning a higher score to utterances of the second language that are not similar to sounds of the language of the non-native speaker.

13. The system of claim 8, wherein normalizing the phonetic difficulty score and the language model score comprises determining a normalized phonetic difficulty score and a normalized language model score such that the mean of the normalized phonetic difficulty score is equal to the mean of the normalized language model score and the standard deviation of the normalized phonetic difficulty score is equal to the standard deviation of the normalized language model score.

14. The system of claim 8, wherein calculating the metric comprises calculating the sum of the normalized phonetic difficulty score and the normalized language model score.

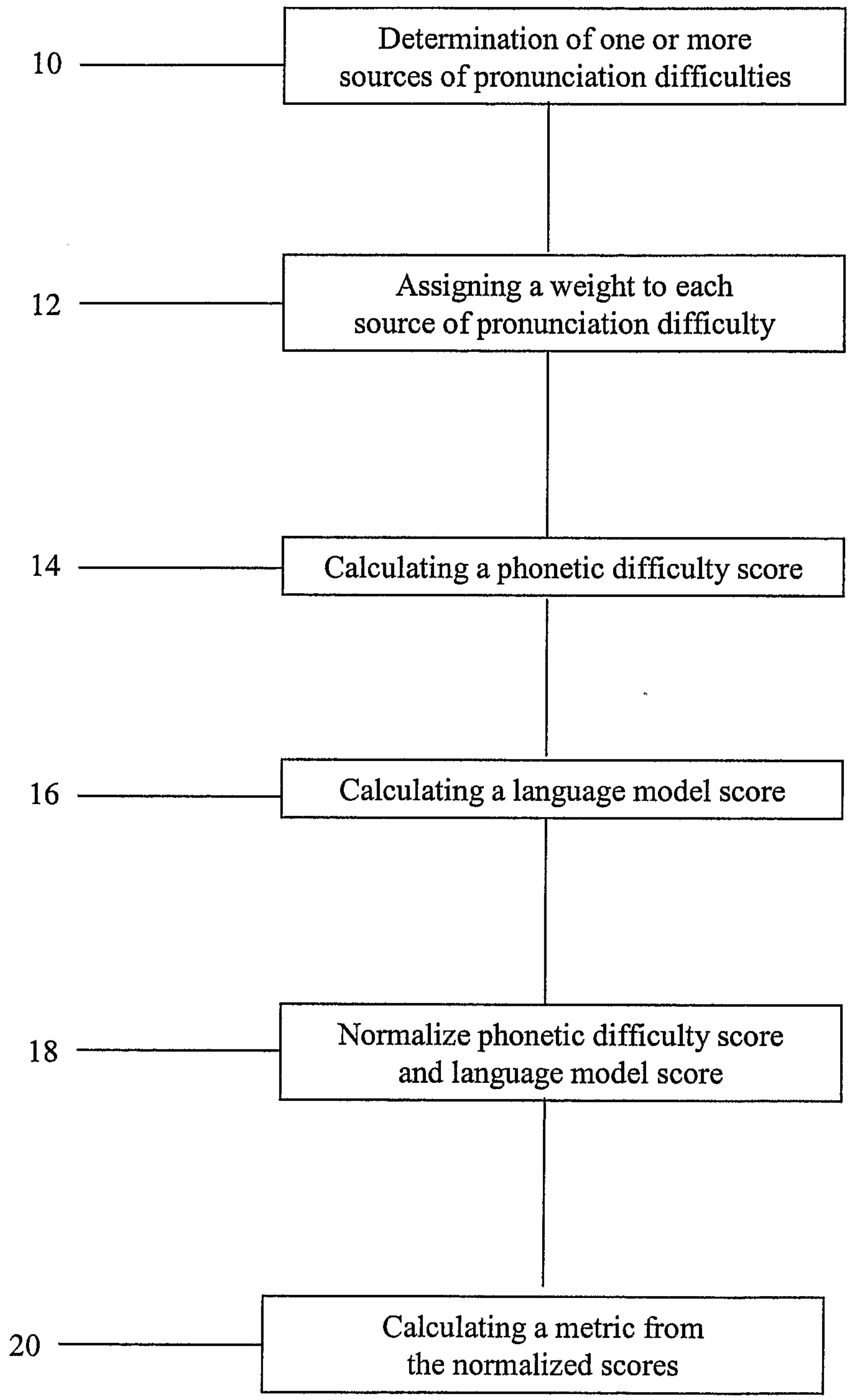
10
Determination of one or more sources of pronunciation difficulties
12
Assigning a weight to each source of pronunciation difficulty
14
Calculating a phonetic difficulty score
16
Calculating a language model score
18
Normalize phonetic difficulty score and language model score
20
Calculating a metric from the normalized scores

FIG. 1

Description	Source	Weight
The sound /r/ (does not contrast with /l/ in Japanese)	/r/	1
The sound /l/ (does not contrast with /r/ in Japanese)	/l/	1
The sound /θ/ (does not exist in Japanese)	/θ/	1
The sound /ð/ (does not exist in Japanese)	/ð/	1
Voiceless stops (Shorter VOT in Japanese)	/p/, /t/, or /k/	0.2
Non-sonorant coda consonant (Japanese syllables are open, or closed only by a nasal)	$/C^{[-sonorant]}]_{\sigma}/$	5
Complex syllable onset (Japanese does not allow consonant clusters	$/[_{\sigma}CC+/$	5

FIG. 2

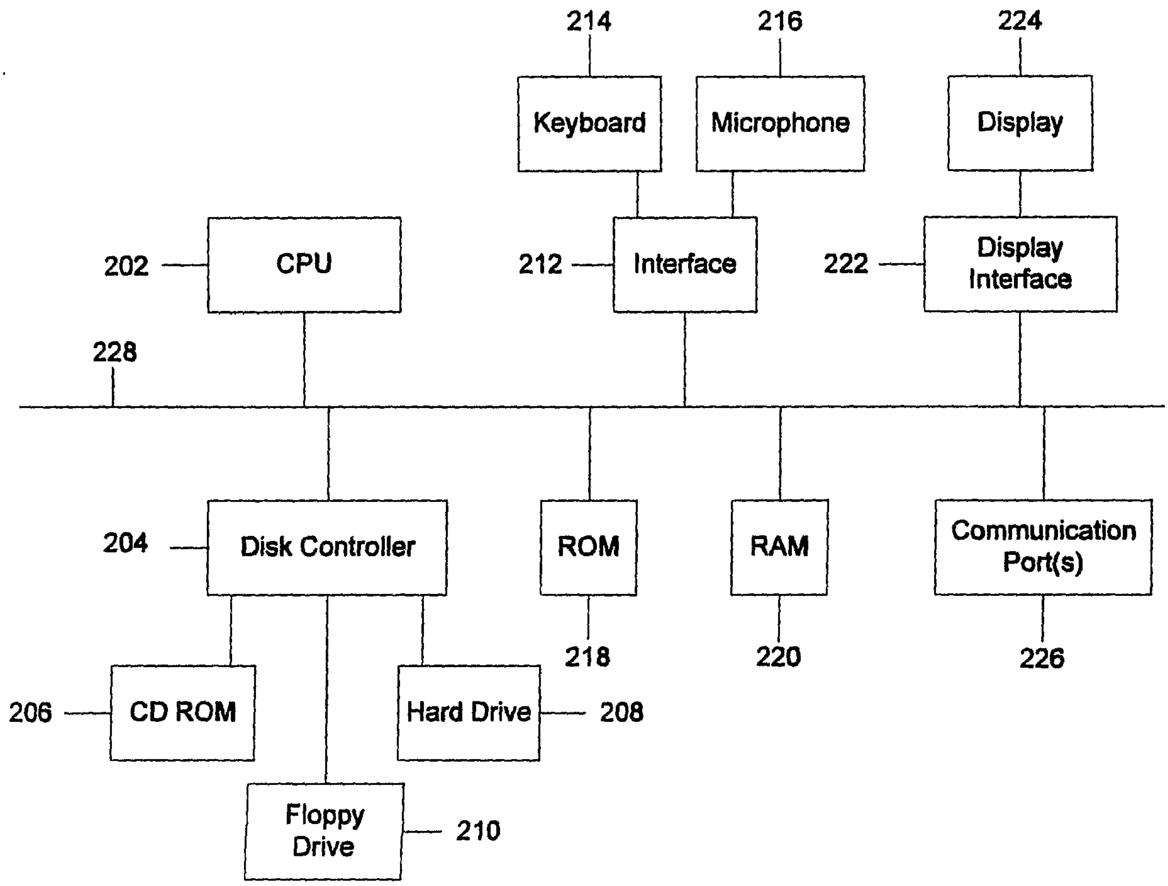
214
216
224
Keyboard
Microphone
Display
202
CPU
212
Interface
222
Display Interface
228
204
Disk Controller
ROM
RAM
Communication Port(s)
218
220
226
206
CD ROM
Hard Drive
208
Floppy Drive
210

FIG. 3