

Please type a plus sign (+) inside this box → +

PTO/SB/05 (11-00)

Approved for use through 10/31/2002. OMB 0651-0032

U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

UTILITY PATENT APPLICATION TRANSMITTAL

Attorney Docket No. 02685-5817

First Inventor Giuseppe RICCARDI

Title METHOD AND SYSTEM FOR BUILDING A PHONOTACTIC MODEL
FOR DOMAIN INDEPENDENT SPEECH RECOGNITION

Express Mail Label No.

(Only for new nonprovisional applications under 37 C.F.R. 1.53(b))

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

1. ☒ Fee Transmittal Form (e.g., PTO/SB/17)
(Submit an original and a duplicate for fee processing)
2. ☐ Applicant claims small entity status.
See 37 CFR 1.27.
3. ☒ Specification [Total Pages 18]
(preferred arrangement set forth below)
 - Descriptive title of the Invention
 - Cross References to Related Applications
 - Statement Regarding Fed sponsored R & D
 - Reference to sequence listing, a table, or a computer program listing appendix
 - Background of the Invention
 - Brief Summary of the Invention
 - Brief Description of the Drawings (if filed)
 - Detailed Description
 - Claim(s)
 - Abstract of the Disclosure
4. ☒ Drawing(s) (35 U.S.C. 113) [Total Sheets 7]
5. Oath or Declaration [Total Pages 3]
 - a. ☒ Newly executed (original or copy)
 - b. ☐ Copy from a prior application (37 CFR 1.63 (d))
(for a continuation/divisional with Box 18 completed)
 - i. ☐ DELETION OF INVENTOR(S)
Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).
6. ☐ Application Data Sheet. See 37 CFR 1.76

ADDRESS TO:

Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

ACCOMPANYING APPLICATIONS PARTS

9. ☒ Assignment Papers (cover sheet & document(s))
10. ☐ 37 C.F.R. §3.73(b) Statement ☐ Power of Attorney
(when there is an assignee)
11. ☐ English Translation Document (if applicable)
12. ☒ Information Disclosure Statement (IDS)/PTO-1449 ☒ Copies of IDS Citations
13. ☐ Preliminary Amendment
14. ☒ Return Receipt Postcard (MPEP 503)
(Should be specifically itemized)
15. ☐ Certified Copy of Priority Document(s)
(if foreign priority is claimed)
16. ☒ Request and Certification under 35 U.S.C. 122
(b)(2)(B)(i). Applicant must attach form PTO/SB/35 or its equivalent.
17. ☐ Other:

18. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment, or in an Application Data Sheet under 37 CFR 1.76:

☐ Continuation ☐ Divisional ☒ Continuation-in-part (CIP) of prior application No: 09/690,721, filed October 18, 2000 and 09/690,903, filed October 18, 2000

☐ Cross Reference to Related Application included in Preliminary Amendment Attached. Note: A cross reference to related application(s) must be filed with the USPTO before four months from the filing date of this continuing application.

For CONTINUATION or DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 5b, is considered a part of the disclosure of the accompanying or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

17. CORRESPONDENCE ADDRESS

☒ Customer Number or Bar Code Label

23838

or ☐ Correspondence address below

(Insert Customer No. or Attach bar code label here)

Name

Ronald E. Prass, Jr., Esq.

Kenyon and Kenyon

Address

1500 K St., N.W., Suite 700

City

Washington, D.C. 20036

Country

U.S.A.

Name (Print/Type)

Ronald E. Prass, Jr.

Registration No. (Attorney/Agent)

42,089

Signature

Date

9/21/01

MICROSOFT CORP.
EXHIBIT 1035

FEE TRANSMITTAL for FY 2001

Patent fees are subject to annual revision.

Complete if Known

Application Number	To Be Assigned
Filing Date	September 21, 2001
First Named Inventor	Giuseppe RICCARDI
Examiner Name	To Be Assigned
Group / Art Unit	To Be Assigned
Attorney Docket No.	02685-5817

TOTAL AMOUNT OF PAYMENT (\$) 1164.00

METHOD OF PAYMENT (check one)		FEE CALCULATION (continued)																																																																																																																																																		
1. ☒ The Commissioner is hereby authorized to charge indicated fees and credit any over payments to: Deposit Account Number: 11-0600 Deposit Account Name: Kenyon & Kenyon ☒ Charge Any Additional Fee Required Under 37 CFR 1.16 and 1.17 ☐ Applicant claims small entity status. See 37 CFR 1.27		3. ADDITIONAL FEES <table border="1"> <thead> <tr> <th>Fee Code</th> <th>Large Entity Fee (\$)</th> <th>Small Entity Fee (\$)</th> <th>Fee Description</th> <th>Fee Paid</th> </tr> </thead> <tbody> <tr><td>105</td><td>130</td><td>205</td><td>65</td><td></td></tr> <tr><td>127</td><td>50</td><td>227</td><td>25</td><td></td></tr> <tr><td>139</td><td>130</td><td>139</td><td>130</td><td></td></tr> <tr><td>147</td><td>2,520</td><td>147</td><td>2,520</td><td></td></tr> <tr><td>112</td><td>920*</td><td>112</td><td>920*</td><td></td></tr> <tr><td>113</td><td>1,840*</td><td>113</td><td>1,840*</td><td></td></tr> <tr><td>115</td><td>110</td><td>215</td><td>55</td><td></td></tr> <tr><td>116</td><td>390</td><td>216</td><td>195</td><td></td></tr> <tr><td>117</td><td>890</td><td>217</td><td>445</td><td></td></tr> <tr><td>118</td><td>1,390</td><td>218</td><td>695</td><td></td></tr> <tr><td>128</td><td>1,890</td><td>228</td><td>945</td><td></td></tr> <tr><td>119</td><td>310</td><td>219</td><td>155</td><td></td></tr> <tr><td>120</td><td>310</td><td>220</td><td>155</td><td></td></tr> <tr><td>121</td><td>270</td><td>221</td><td>135</td><td></td></tr> <tr><td>138</td><td>1,510</td><td>138</td><td>1,510</td><td></td></tr> <tr><td>140</td><td>110</td><td>240</td><td>55</td><td></td></tr> <tr><td>141</td><td>1,240</td><td>241</td><td>620</td><td></td></tr> <tr><td>142</td><td>1,240</td><td>242</td><td>620</td><td></td></tr> <tr><td>143</td><td>440</td><td>243</td><td>220</td><td></td></tr> <tr><td>144</td><td>600</td><td>244</td><td>300</td><td></td></tr> <tr><td>122</td><td>130</td><td>122</td><td>130</td><td></td></tr> <tr><td>123</td><td>130</td><td>123</td><td>130</td><td></td></tr> <tr><td>126</td><td>180</td><td>126</td><td>180</td><td></td></tr> <tr><td>581</td><td>40</td><td>581</td><td>40</td><td>40.00</td></tr> <tr><td>146</td><td>710</td><td>246</td><td>355</td><td></td></tr> <tr><td>149</td><td>710</td><td>249</td><td>355</td><td></td></tr> <tr><td>179</td><td>710</td><td>279</td><td>355</td><td></td></tr> <tr><td>169</td><td>900</td><td>169</td><td>900</td><td></td></tr> </tbody> </table>		Fee Code	Large Entity Fee (\$)	Small Entity Fee (\$)	Fee Description	Fee Paid	105	130	205	65		127	50	227	25		139	130	139	130		147	2,520	147	2,520		112	920*	112	920*		113	1,840*	113	1,840*		115	110	215	55		116	390	216	195		117	890	217	445		118	1,390	218	695		128	1,890	228	945		119	310	219	155		120	310	220	155		121	270	221	135		138	1,510	138	1,510		140	110	240	55		141	1,240	241	620		142	1,240	242	620		143	440	243	220		144	600	244	300		122	130	122	130		123	130	123	130		126	180	126	180		581	40	581	40	40.00	146	710	246	355		149	710	249	355		179	710	279	355		169	900	169	900	
Fee Code	Large Entity Fee (\$)	Small Entity Fee (\$)	Fee Description	Fee Paid																																																																																																																																																
105	130	205	65																																																																																																																																																	
127	50	227	25																																																																																																																																																	
139	130	139	130																																																																																																																																																	
147	2,520	147	2,520																																																																																																																																																	
112	920*	112	920*																																																																																																																																																	
113	1,840*	113	1,840*																																																																																																																																																	
115	110	215	55																																																																																																																																																	
116	390	216	195																																																																																																																																																	
117	890	217	445																																																																																																																																																	
118	1,390	218	695																																																																																																																																																	
128	1,890	228	945																																																																																																																																																	
119	310	219	155																																																																																																																																																	
120	310	220	155																																																																																																																																																	
121	270	221	135																																																																																																																																																	
138	1,510	138	1,510																																																																																																																																																	
140	110	240	55																																																																																																																																																	
141	1,240	241	620																																																																																																																																																	
142	1,240	242	620																																																																																																																																																	
143	440	243	220																																																																																																																																																	
144	600	244	300																																																																																																																																																	
122	130	122	130																																																																																																																																																	
123	130	123	130																																																																																																																																																	
126	180	126	180																																																																																																																																																	
581	40	581	40	40.00																																																																																																																																																
146	710	246	355																																																																																																																																																	
149	710	249	355																																																																																																																																																	
179	710	279	355																																																																																																																																																	
169	900	169	900																																																																																																																																																	
2. ☐ Payment Enclosed: ☐ Check ☐ Credit card ☐ Money Order ☐ Other		Other fee (specify) _____																																																																																																																																																		
FEE CALCULATION 1. BASIC FILING FEE <table border="1"> <thead> <tr> <th>Large Fee Code</th> <th>Entity Fee (\$)</th> <th>Small Fee Code</th> <th>Entity Fee (\$)</th> <th>Fee Description</th> <th>Fee Paid</th> </tr> </thead> <tbody> <tr><td>101</td><td>710</td><td>201</td><td>355</td><td>Utility filing fee</td><td>710.00</td></tr> <tr><td>106</td><td>320</td><td>206</td><td>160</td><td>Design filing fee</td><td></td></tr> <tr><td>107</td><td>490</td><td>207</td><td>245</td><td>Plant filing fee</td><td></td></tr> <tr><td>108</td><td>710</td><td>208</td><td>355</td><td>Reissue filing fee</td><td></td></tr> <tr><td>114</td><td>150</td><td>214</td><td>75</td><td>Provisional filing fee</td><td></td></tr> </tbody> </table> SUBTOTAL (1) (\$) 710.00		Large Fee Code	Entity Fee (\$)	Small Fee Code	Entity Fee (\$)	Fee Description	Fee Paid	101	710	201	355	Utility filing fee	710.00	106	320	206	160	Design filing fee		107	490	207	245	Plant filing fee		108	710	208	355	Reissue filing fee		114	150	214	75	Provisional filing fee		2. EXTRA CLAIM FEES <table border="1"> <thead> <tr> <th>Total Claims</th> <th>Independent Claims</th> <th>Multiple Dependent</th> <th>Extra Claims</th> <th>Fee from below</th> <th>Fee Paid</th> </tr> </thead> <tbody> <tr> <td>43</td> <td>-20</td> <td></td> <td>23</td> <td>18.00</td> <td>414.00</td> </tr> <tr> <td>3</td> <td>-3</td> <td></td> <td>0</td> <td></td> <td>0</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td>0</td> </tr> </tbody> </table> Large Fee Code Entity Fee (\$) <table border="1"> <thead> <tr> <th>Large Fee Code</th> <th>Entity Fee (\$)</th> <th>Small Fee Code</th> <th>Entity Fee (\$)</th> <th>Fee Description</th> <th>Fee Paid</th> </tr> </thead> <tbody> <tr><td>103</td><td>18</td><td>203</td><td>9</td><td>Claims in excess of 20</td><td></td></tr> <tr><td>102</td><td>80</td><td>202</td><td>40</td><td>Independent claims in excess of 3</td><td></td></tr> <tr><td>104</td><td>270</td><td>204</td><td>135</td><td>Multiple dependent claim, if not paid</td><td></td></tr> <tr><td>109</td><td>80</td><td>209</td><td>40</td><td>** Reissue independent claims over original patent</td><td></td></tr> <tr><td>110</td><td>18</td><td>210</td><td>9</td><td>** Reissue claims in excess of 20 and over original patent</td><td></td></tr> </tbody> </table> SUBTOTAL (2) (\$) 414.00		Total Claims	Independent Claims	Multiple Dependent	Extra Claims	Fee from below	Fee Paid	43	-20		23	18.00	414.00	3	-3		0		0						0	Large Fee Code	Entity Fee (\$)	Small Fee Code	Entity Fee (\$)	Fee Description	Fee Paid	103	18	203	9	Claims in excess of 20		102	80	202	40	Independent claims in excess of 3		104	270	204	135	Multiple dependent claim, if not paid		109	80	209	40	** Reissue independent claims over original patent		110	18	210	9	** Reissue claims in excess of 20 and over original patent																																																		
Large Fee Code	Entity Fee (\$)	Small Fee Code	Entity Fee (\$)	Fee Description	Fee Paid																																																																																																																																															
101	710	201	355	Utility filing fee	710.00																																																																																																																																															
106	320	206	160	Design filing fee																																																																																																																																																
107	490	207	245	Plant filing fee																																																																																																																																																
108	710	208	355	Reissue filing fee																																																																																																																																																
114	150	214	75	Provisional filing fee																																																																																																																																																
Total Claims	Independent Claims	Multiple Dependent	Extra Claims	Fee from below	Fee Paid																																																																																																																																															
43	-20		23	18.00	414.00																																																																																																																																															
3	-3		0		0																																																																																																																																															
					0																																																																																																																																															
Large Fee Code	Entity Fee (\$)	Small Fee Code	Entity Fee (\$)	Fee Description	Fee Paid																																																																																																																																															
103	18	203	9	Claims in excess of 20																																																																																																																																																
102	80	202	40	Independent claims in excess of 3																																																																																																																																																
104	270	204	135	Multiple dependent claim, if not paid																																																																																																																																																
109	80	209	40	** Reissue independent claims over original patent																																																																																																																																																
110	18	210	9	** Reissue claims in excess of 20 and over original patent																																																																																																																																																
SUBTOTAL (3) (\$) 40.00		*Reduced by Basic Filing Fee Paid																																																																																																																																																		

SUBMITTED BY		Complete (if applicable)	
Name (Print/Type)	Ronald E. Prass, Jr., Esq.	Registration No. Attorney/Agent)	42,089
Signature		Telephone	202-220-4207
		Date	September 21, 2001

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

REQUEST AND CERTIFICATION UNDER 35 U.S.C. 122(b)(2)(B)(i)	First Named Inventor		Giuseppe RICCARDI
	Title	METHOD AND SYSTEM FOR BUILDING A PHONOTACTIC MODEL FOR DOMAIN INDEPENDENT SPEECH RECOGNITION	
	Atty Docket Number		02685-5817

I hereby certify that the invention disclosed in the attached application **has not and will not be** the subject of an application filed in another country, or under a multilateral agreement, that requires publication at eighteen months after filing. I hereby request that the attached application not be published under 35 U.S.C. 122(b).

SEPTEMBER 21, 2001

Date


Signature

Ronald E. Prass, Jr., Esq.,

Typed or printed name

This request must be signed in compliance with 37 CFR 1.33(b) and submitted with the application **upon filing**.

Applicant may rescind this nonpublication request at any time. If applicant rescinds a request that an application not be published under 35 U.S.C. 122(b), the application will be scheduled for publication at eighteen months from the earliest claimed filing date for which a benefit is claimed.

If applicant subsequently files an application directed to the invention disclosed in the attached application in another country, or under a multilateral international agreement, that requires publication of applications eighteen months after filing, the applicant **must** notify the United States Patent and Trademark Office of such filing within forty-five (45) days after the date of the filing of such foreign or international application. **Failure to do so will result in abandonment of this application (35 U.S.C. 122(b)(2)(B)(iii)).**

Burden Hour Statement: This collection of information is required by 37 CFR 1.213(a). The information is used by the public to request that an application not be published under 35 U.S.C. 122(b) (and the PTO to process that request). Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This form is estimated to take 6 minutes to complete. This time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

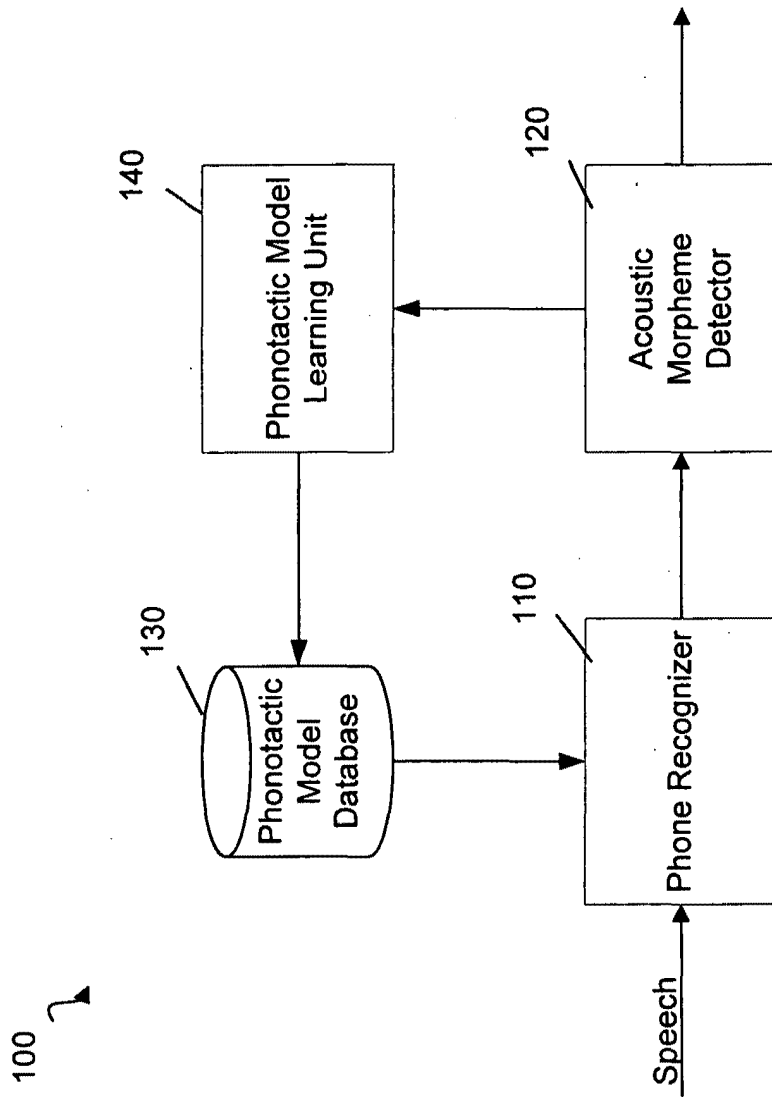


Fig. 1

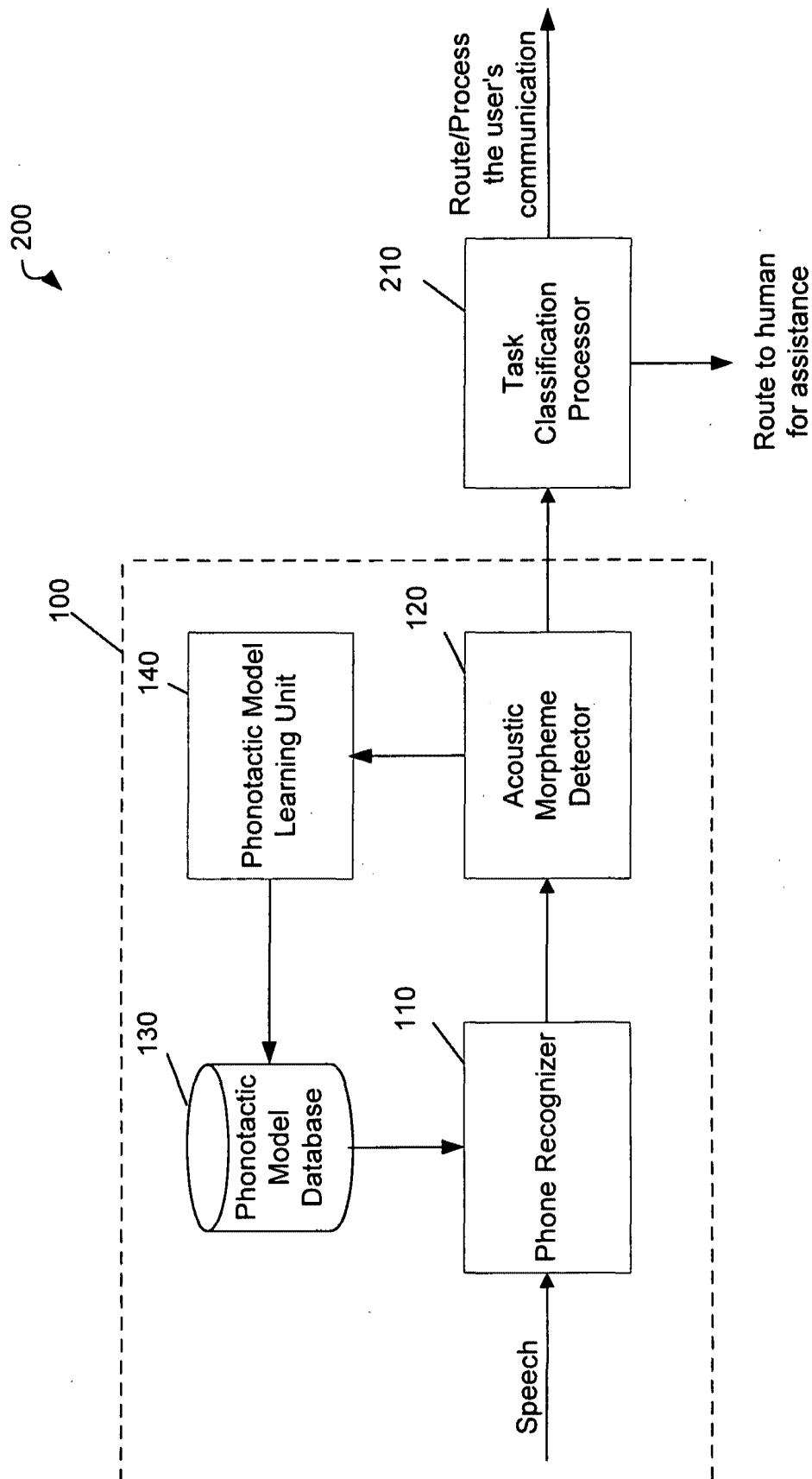


Fig. 2

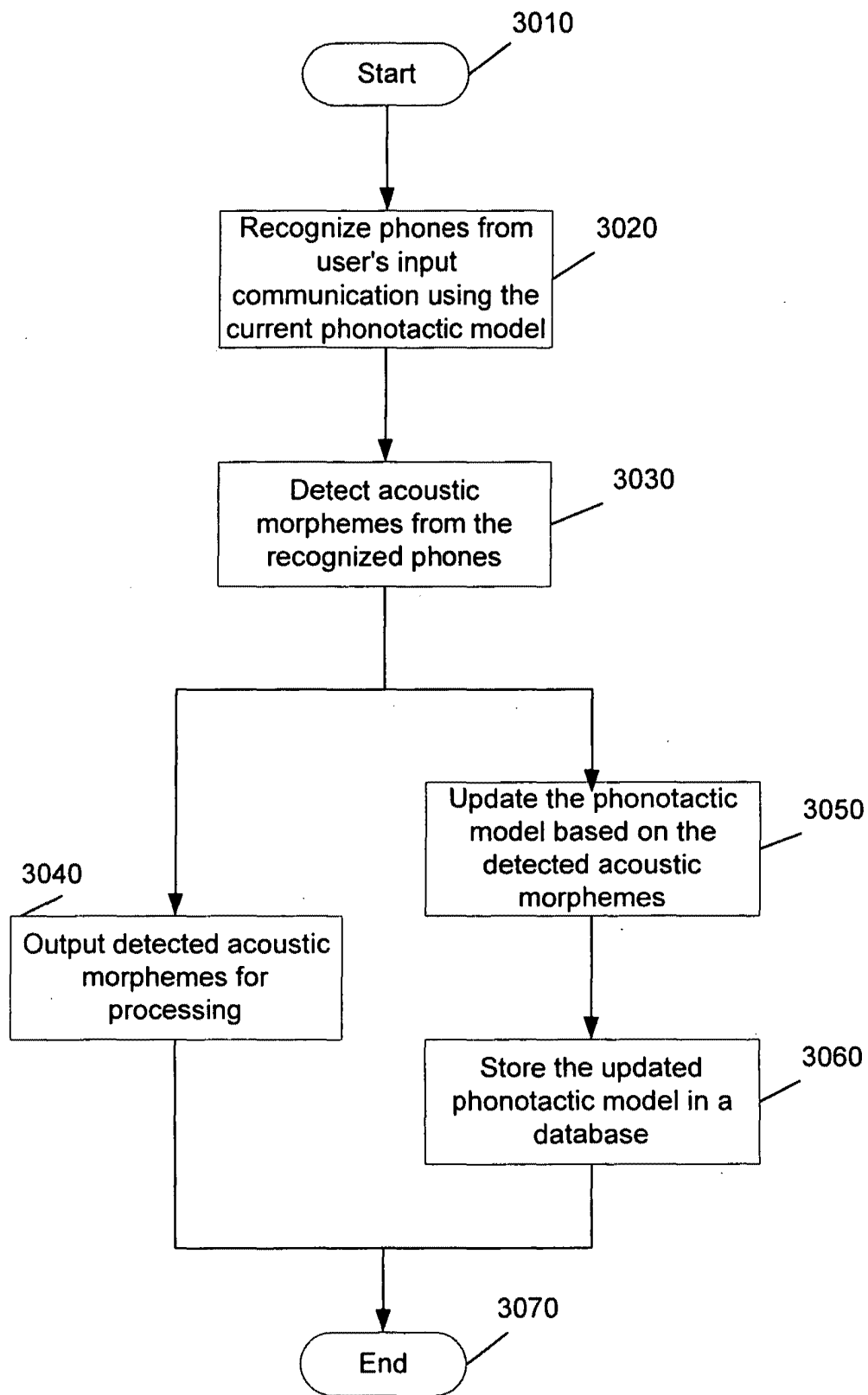


Fig. 3

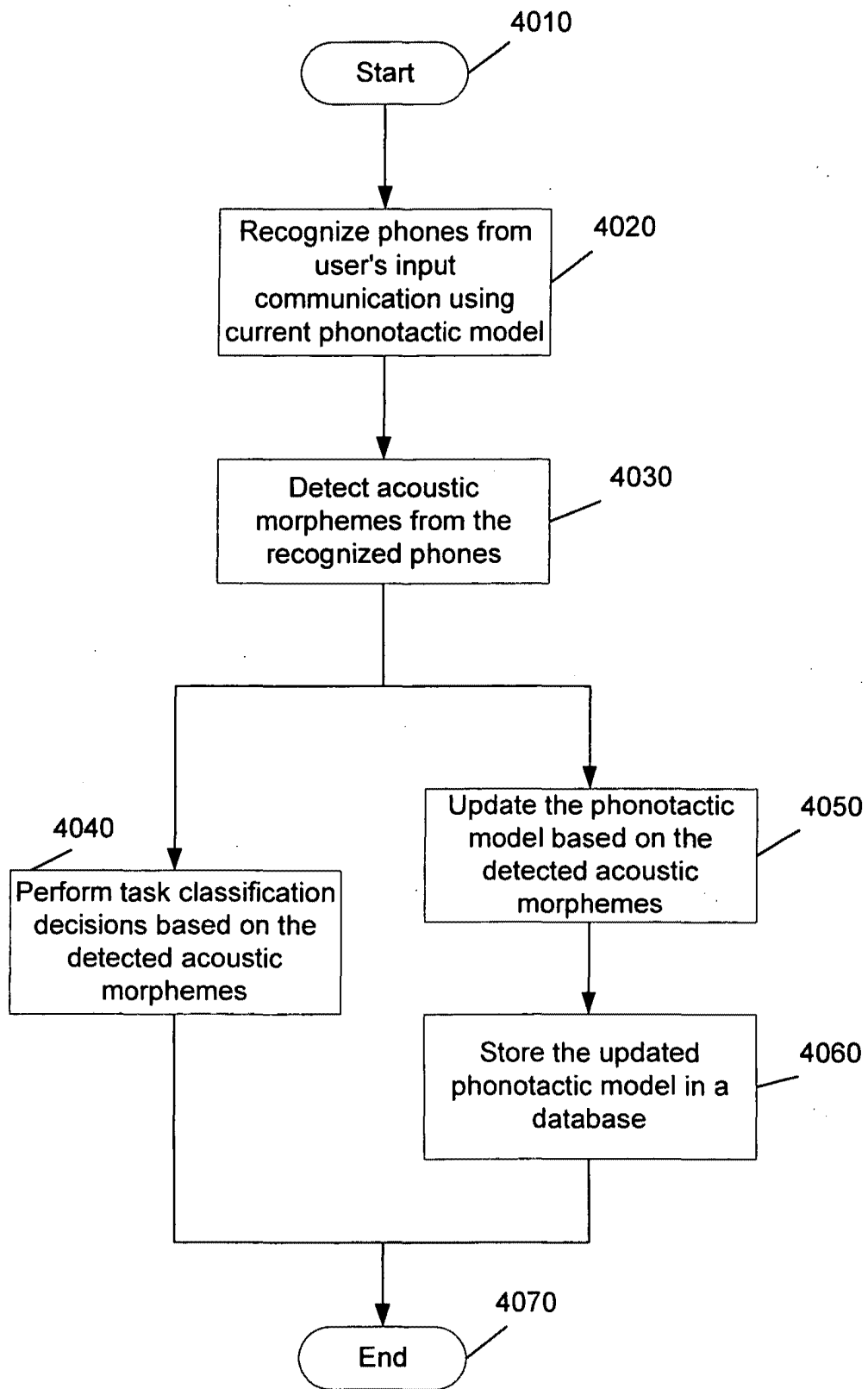


Fig. 4

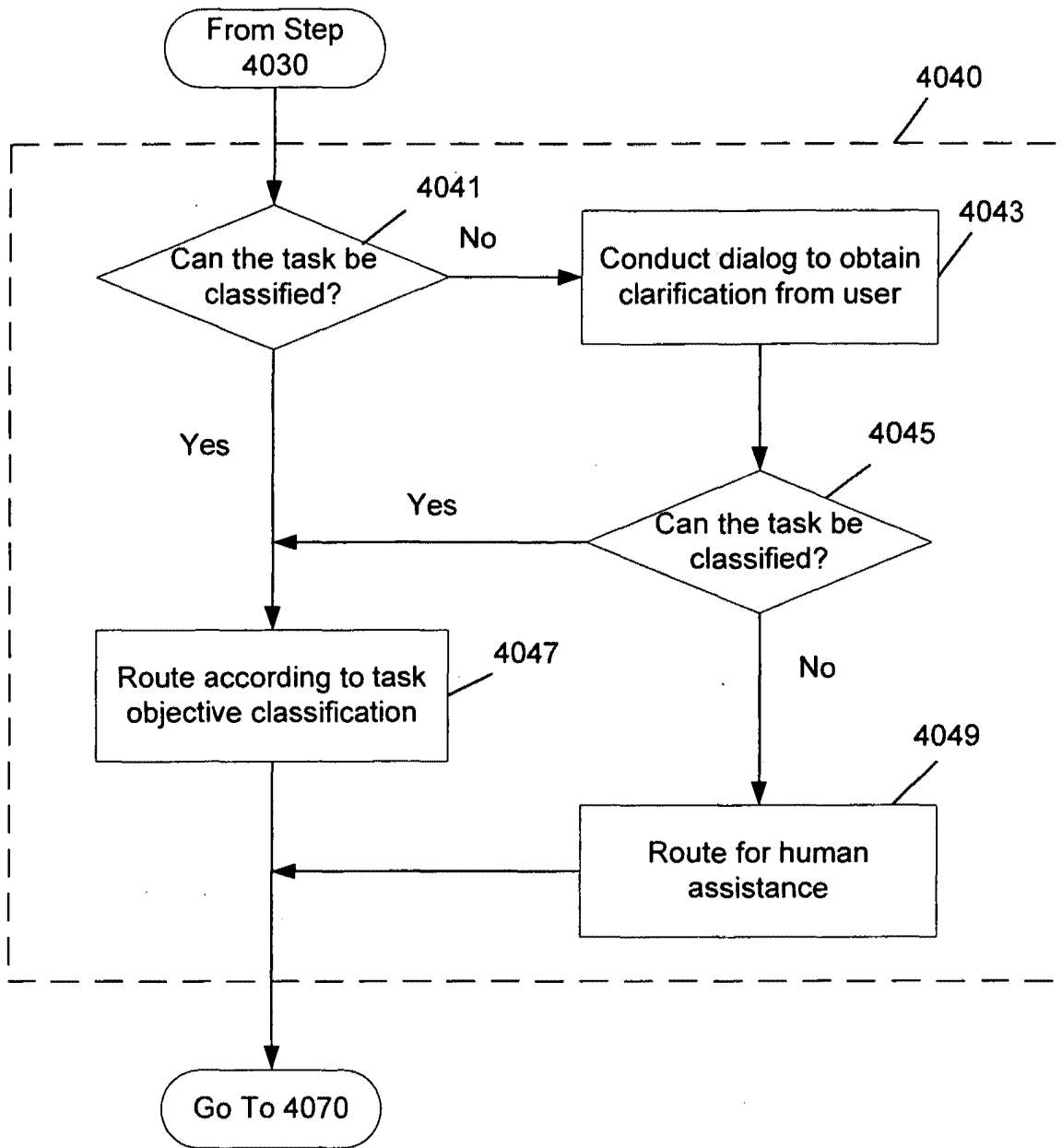


Fig. 5

<u>MI</u>	<u>WMI</u>				
7.75	0.0611	K a o l			Call
10.00	0.0590	n a h m B e r			Number
9.15	0.0421	f a o r			Four
14.37	0.0280	y e h s			Yes
8.31	0.0370	w a h n			One
10.00	0.0333	s e h v a x n			Seven
26.00	0.0277	K a o l i h n g K a a r D			Calling
					Card
3.63	0.0276	T u w			Two/To
6.12	0.0227	L a y K			Like

Fig. 6

A scatter plot showing the relationship between F1 score (Y-axis) and F2 score (X-axis) for two datasets: On-Line Learning VS HMIHY text (represented by solid circles) and SWBD VS HMIHY text (represented by open circles). The plot uses a logarithmic scale for both axes, ranging from 10^{-8} to 10^0 on the X-axis and 10^{-5} to 10^0 on the Y-axis. A dashed diagonal line represents the identity line where F1 = F2. The data points are clustered around this line, indicating that the F1 and F2 scores are generally similar for both datasets.

Fig. 8

5

METHOD AND SYSTEM FOR BUILDING A PHONOTACTIC MODEL FOR DOMAIN INDEPENDENT SPEECH RECOGNITION

10

Claim for Priority/Cross Reference to Related Applications

[0001] This non-provisional application claims the benefit of U.S. Provisional Patent Application No. 60/235,861, filed September 27, 2000, which is incorporated by reference in its entirety. This non-provisional application is also a continuation-in-part of co-pending U.S. Patent Applications Nos. 09/690,721 and 09/690,903, both filed October 18, 2000, which claim the benefit of U.S. Provisional Patent Application No. 60/163,838, filed November 5, 1999, which are also incorporated herein by reference in their entireties.

Technical Field

[0002] The invention relates to automated systems for communication recognition and understanding.

Background Of The Invention

[0003] Conventional methods for constructing spoken language systems involve training speech and language models prior to operation by transcribing speech input and finding existing patterns. This speech is manually transcribed and each utterance is then semantically labeled. The resultant database is exploited to train stochastic language models for recognition and understanding. These models are further adapted for different dialog states. Examples of such methods are shown in U.S. Patent Nos. 5,675,707, 5,860,063, 6,044,337,

6,173,261, 6,021,384 and 6,192,110, each of which is incorporated by reference herein in its entirety.

[0004] This transcription and labeling process is a major bottleneck in new application development and refinement of existing ones. For incremental
5 training of a deployed natural spoken dialog system, current technology would potentially require transcribing millions of transactions. This process is both time-consuming and prohibitively expensive.

Summary Of The Invention

[0005] The invention concerns a method and corresponding system for
10 building a phonotactic model for domain independent speech recognition. The method may include recognizing phones from a user's input communication using a current phonotactic model, detecting morphemes (acoustic and/or non-acoustic) from the recognized phones, and outputting the detected morphemes for processing. The method also updates the phonotactic model with the
15 detected morphemes and stores the new model in a database for use by the system during the next user interaction. The method may also include making task-type classification decisions based on the detected morphemes from the user's input communication.

Brief Description of the Drawings

[0006] The invention is described in detail with reference to the following drawings wherein like numerals reference like elements, and wherein:

[0007] Fig. 1 is a block diagram of an exemplary speech recognition system of the invention;

[0008] Fig. 2 is a block diagram of an exemplary speech recognition
25 system of the invention used in conjunction with a task classification system;

[0009] Fig. 3 is a flowchart illustrating an exemplary phonotactic model updating process;

[0010] Fig. 4 is a flowchart illustrating an exemplary phonotactic model updating process in a task classification system;

[0014] Fig. 8 is a graph illustrating the on-line oracle experimental results as compared to the estimates from text transcriptions.

10 [0015] Traditional ASR techniques require the domain-specific knowledge
of acoustic, lexicon data and more importantly the word probability distributions.
However, this invention concerns a method for unsupervised learning of acoustic
and lexical units from out-of-domain speech data. The new lexical units are used
for fast adaptation of language model probabilities to a new domain. Starting
15 from a lexicon and language model, the relevant language statistics of task
classification systems may be learned.

[0017] The conventional large vocabulary speech recognition framework requires acoustic and language model to be trained on domain-specific data. This data is usually collected through human-to-machine interaction protocols and speech utterances are transcribed for the purpose of acoustic and language model training. These models perform poorly in out-of-domain conditions and are not suitable for on-line learning of language. This is true despite the fact that

large vocabulary lexicons can have millions of words and reduce the out-of vocabulary rate to zero. However, the main reason for such poor performance is the mismatch of the language model probabilities.

[0018] An alternative approach to large vocabulary recognition is to model phone sequences. There has been active research on this topic and high phone recognition rates have been achieved for restricted domains and language.

While it is widely accepted that phone recognition performance is inferior to word-based large vocabulary speech recognition tasks (where large databases are available to train language models), phone recognition allows a system to perform task-independent speech recognition. For these reasons, task-independent speech recognizers should combine the accuracy of a word-based system and the acoustic event granularity of a phone-based system.

[0019] Baseline approaches to the teaching of speech recognition systems are found in U.S. Patents Nos. 5,675,707 and 5,860,063, 6,173,261, and 6,192,110, which have been incorporated above herein by reference in their entireties.

[0020] While the morphemes may be non-acoustic (i.e., made up of non-verbal sub-morphemes such as tablet strokes, gestures, body movements, etc.), for ease of discussion, the systems and methods illustrated in the drawings and discussed in the below concern only acoustic morphemes. Consequently, the invention should not be limited to just acoustic morphemes and should encompass the utilization of any sub-units of any known or future method of communication for the purposes of recognition and understanding.

[0021] Furthermore, while the terms "speech", "phrase" and "utterance", used throughout the description below, may connote only spoken language, it is important to note in the context of this invention, "speech", "phrase" and "utterance" may include verbal and/or non-verbal sub-units (or sub-morphemes). Therefore, "speech", "phrase" and "utterance" may comprise non-verbal sub-units, verbal sub-units or a combination of verbal and non-verbal sub-units within the spirit and scope of this invention.

[0022] Non-verbal speech may include but are not limited to gestures, body movements, head movements, non-responses, text, keyboard entries, keypad entries, mouse clicks, DTMF codes, pointers, stylus, cable set-top box entries, graphical user interface entries and touchscreen entries, or a combination thereof. Multimodal information is received using multiple channels (i.e., aural, visual, etc.). The user's input communication may also be derived from the verbal and non-verbal speech and the user's or the machine's environment. Basically, any manner of communication falls within the intended scope of the invention. However, for ease of discussion, we will focus on verbal speech in the examples and embodiments set forth below.

[0023] Fig. 1 is an exemplary block diagram of an exemplary speech recognition and understanding system 100 that utilizes acoustic morphemes. The speech recognition and understanding system 100 may include a phone recognizer 110, an acoustic morpheme detector 120, a phonotactic model learning unit 140, and a phonotactic database 130.

[0024] In the speech recognition and understanding system 100, the phone recognizer 110 may receive an input communications, such as speech, from a user, customer, etc., in the form of verbal and/or non-verbal speech. As discussed above, the input speech may, of course, be expressed in verbal speech, non-verbal speech, multimodal forms, or using a mix of verbal and non-verbal speech.

[0025] The phone recognizer 110 is the task-independent component of the speech recognition and understanding system 100 which transcribes speech utterances from one domain and is trained on another domain. Prior to being integrated in the system 100, the phone recognizer 110 may be initially trained from an off-the-shelf database, for example. The database may be generated from recordings of users talking with human agents, responding to the prompt "AT&T, How may I help you?" (HMIHY). The characteristics of this data and early experiments are detailed in U.S. Patent No. 5,675,707, for example, which has been incorporated herein by reference in its entirety. In any event, the phone recognizer 110 may be of any known design and performs the function of

recognizing, or spotting, the existence of one or more phone in the user's input speech.

[0026] In an embodiment for recognizing non-acoustic morphemes, the phone recognizer 110 may be replaced in the figure by a sub-morpheme

5 recognizer. The sub-morpheme recognizer would operate similar to the phone recognizer 110, but it would receive raw non-acoustic or a mixture of acoustic and non-acoustic speech from the user. Again, although the invention encompasses non-acoustic speech, for ease of discussion, only acoustic morphemes will be discussed in reference to the figures.

10 [0027] The acoustic morpheme detector 120 detects the acoustic morphemes present in the recognized phones from user's input request. The acoustic morphemes detected by acoustic morpheme detector 120 may be used for processing by any number of automated systems known to those of ordinary skill in the art. For example, the acoustic morphemes detected by the acoustic
15 morpheme detector 120 may include a large number of verbal and non-verbal speech fragments or sub-morphemes (illustrated as phone-phrases for ease of discussion), each of which may be related to one of a predetermined set of task objectives. In this respect, each of the acoustic morphemes may be labeled with its associated task objective. Fig.2 shows an example of one such task
20 classification system 200, as discussed below.

[0028] The acoustic morphemes detected by the acoustic morpheme detector 120 are also provided to the phonotactic model learning unit 140 for building a new phonotactic model. The phonotactic model learning unit 140 generates a new phonotactic model based on the detected acoustic morphemes
25 found in the recognized phones from the user's input speech. The newly generated phonotactic model is then stored in the phonotactic model database 130 for use by the phone recognizer 110.

[0029] Fig. 2 shows the use of detected acoustic morphemes resulting from the automatically generated phonotactic model in a task classification
30 system 200. The task classification system 200 may include a task classification processor 210 along with the same elements as the speech recognition and

understanding system 100. The task classification processor 210 may apply a confidence function, based on the probabilistic relation between the recognized acoustic morphemes and selected task objectives, and makes a decision either to implement a particular task objective, or makes a determination that no decision is likely in which case the user may be defaulted to a human or automated system for assistance. Within this framework, on-line estimates of word or morpheme probabilities are learned for task-independent speech recognition. In fact, a classifier that operates on words, morphemes (acoustic and non-acoustic), and/or sub-morphemes, etc., may be used within the spirit and scope of the invention.

[0030] An exemplary process of the invention will now be described in Figs. 3-5 with reference to the exemplary structure shown in Figs. 1 and 2. Fig. 3 illustrates a possible process of automatically building a phonotactic model using acoustic morphemes for use in the speech recognition and understanding system 100. The process begins at step 3010 and proceeds to step 3020 where the phone recognizer 110 receives raw speech from a user and recognizes phones in the speech using the current phonotactic model stored in the phonotactic model database 130. In the phonotactic model, each word in the speech corpus is replaced with its most likely pronunciation and trained n-gram model based on the Variable N-gram Stochastic Automaton (VNSA).

[0031] The acoustic morpheme detector 120 may detect phone sequences by computing information theoretic measures, such as the weighted mutual information $I_W(x,y) = P(x,y)MI(x,y) = P(x,y)\log P(x,y)/P(x)P(y)$, where x and y are two phone sequences. Weighted mutual information compensates for probability estimates biases introduced by rare events in the mutual information, $MI(x,y)$. Alternative and computationally expensive methods for computing acoustic morphemes are entropy-minimizing units and multi-gram approaches. It can be shown both theoretically and experimentally that the n-best I_W is a close approximation of the exact entropy minimization problem. Such methods are shown in U.S. Patent No. 6,021,384, which has been incorporated by reference above.

[0032] Excerpts from the detected acoustic morphemes are shown in Fig. 6. The first and second column contain the I_w and M/I values, the third and fourth are the corresponding acoustic morpheme sequence and its most likely word sequence. The phonotactic model learning unit 140 may for example, incorporate the top N ($N = 500$) acoustic morphemes in the stochastic phonotactic models which takes advantage of the long-spanning acoustic morphemes.

[0033] After the acoustic morphemes have been detected, the process proceeds on a dual track of processing and training. In the processing track, in step 3040, the acoustic morphemes are output for processing. One of the possible processes for detected acoustic morphemes may be to map them into word sequences using a lexical access unit. In general, this process is not needed if the semantic units are not based on words. However, word-based language models are the natural candidates for cross-domain modeling, given the large availability of transcribed speech databases and the effectiveness word-based large vocabulary speech recognition.

[0034] Acoustic morphemes can be mapped to zero, one or many word sequences, for a given baseform lexicon. A conservative strategy to lexical access may be taken which considers only the exact match between baseforms in the lexicon and acoustic morphemes. This conservative strategy is dictated by the need to learn the most relevant lexical features of the domain with high precision and high rejection. These lexical features may be used to adapt on-line an off-the-shelf large vocabulary language model. The acquisition algorithm discussed above has selected 500 units that were matched against their most likely word mapping drawn from the lexicon. For example, in Fig. 6, a list of words and phrases was compiled. While the list of lexical items is short, due to the conservative strategy, they cover 40% of the probability mass of the speech corpus.

[0035] Alternatively, the detected acoustic morphemes may be used by a task classification system 200 from Fig. 2, as depicted in the exemplary flowchart shown in Fig. 4. Fig. 4 includes steps 4010, 4020, 4030, 4050, 4060 and 4070

that correspond to steps 3010, 3020, 3030, 3050, 3060 and 3070 of Fig. 3. However, Fig. 4 also includes step 4040 where the task classification processor 210 makes task classification decisions based on the detected acoustic morphemes.

5 [0036] As an example of processes that may be performed as part of a possible task classification process using acoustic morphemes, Fig. 5 is provided. In Fig. 5, from step 4030 where acoustic morphemes are detected from recognized phones from a user's input communication, the process proceeds to step 4040 and sub-steps 4041-4049 in which the task classification
10 processor 210 performs task classifications based on the detected acoustic morphemes. The task classification processor 210 may apply a confidence function based on the probabilistic relation between the recognized acoustic morphemes and selected task objectives, for example.

[0037] In step 4041, the task classification processor 210 determines
15 whether a task can be classified based on the detected acoustic morpheme. If the task can be classified, in step 4047, the task classification processor 210 routes the user/customer according to the classified task objective. The process then goes to step 4070 and ends.

[0038] If the task cannot be classified in step 4041 (i.e. a low confidence
20 level has been generated), in step 4043, a dialog module (located internally or externally) the task classification processor 210 conducts dialog with the user/customer to obtain clarification of the task objective. After dialog has been conducted with the user/customer, in step 4045, the task classification processor 210 determines whether the task can now be classified based on the additional
25 dialog. If the task can be classified, the process proceeds to step 4047 and the user/customer is routed in accordance with the classified task objective and the process ends at step 4070. However, if task can still not be classified, in step 4049, the user/customer is routed to a human for assistance and then the process goes to step 4070 and ends.

30 [0039] Referring back to Figs. 3, after the acoustic morphemes have been detected in step 3030, they may be used in a processed for building a

phonotactic model using the phonotactic model learning unit 140 for ultimate use by the phone recognizer 110 in recognizing phone sequences in the user's next utterance. In building the phonotactic model, the phonotactic learning unit 140 transforms a prior probability distribution from one domain to another without speech transcriptions (on-line adaptation). This is the primary constraint of an adaptive spoken dialog system, which can react and recover to novel events is the syntactic and semantic input channel. Most of the literature in language model adaptation is for batch processing (with speech transcriptions) and assumes that the probability space is the same both for the source and target domain, namely the word sequence space.

[0040] The building of the phonotactic model by the phohtactic model learning unit 140 according to the invention involves at least two stochastic modeling processes. The first is to map phone sequence statistics into word statistics. If the most likely pronunciation of the word *collect* is *K ae l eh K t*, then $P(\text{collect}) \neq P(K ae l eh K t)$. In general terms:

$$P_{on-line}(\omega) = \sum_{f_i \in B_\omega} P(f_i)$$

where f_i is a baseform of the word ω drawn from the set B_ω . It is assumed from

here forward that $P_{on-line}(\omega) \approx P(f_i)$, where f_i is the most likely pronunciation.

[0041] Fig. 7 is a scatter plot $(\log P_{HMIHY}(\omega), \log P_{on-line}(\omega))$ (asterisk) and $(\log P_{HMIHY}(\omega), \log P_{switchboard}(\omega))$ (circle), where both $P_{HMIHY}(\omega)$ and $P_{switchboard}(\omega)$ have been estimated from the hand-labeled speech transcriptions. Most word probability estimates lie along the diagonal (dashed line) and closely approximate the true empirical distribution $P_{HMIHY}(\omega)$. In order to get an upper bound on the on-line probability estimates the phone sequence with lowest string edit distance from the reference phone transcription was computed. Thus, we have compiled the oracle phone transcription training set and estimated $P_{oracle online}(\omega)$.

[0042] Fig. 8 is a scatter plot of the log-probabilities for the on-line oracle estimates. From Fig. 8 we see that the probabilities of the selected words and

phrases have been estimated with high precision with respect to the best possible guess (oracle).

[0043] The second problem is to transform the word probability vector of the prior distribution using the word statistics learned on-line. Given the small lexical coverage and large probability mass of the acquired features the two distributions have been merged with following scheme:

$$P_{target}(\omega) = \begin{cases} P_{on-line}(\omega) & w \in S \\ \frac{1-\alpha}{\beta} P_{switchboard}(\omega) & w \in S \end{cases}$$

where:

$$\alpha = \sum_{w \in S} P_{on-line}(\omega), \beta = \sum_{w \in S, \sum_{w \in V} P(\omega)}$$

and S is the set of selected words. This model has been tested to measure the perplexity of the HMIHY test set perplexity for a unigram model. The unigram perplexity on the HMIHY test set, in matched language model condition ($P_{HMIHY}(\omega)$) is 128.3. The test perplexity using the switchboard language model ($P_{switchboard}(\omega)$) is 715.9. The test set perplexity was tested with the new word probability $P_{target}(\omega)$ and achieved a 39.4% relative improvement (433.9) with respect to the switchboard baseline.

[0044] Once the new phonotactic model is built, in step 3060 it is stored in the phonotactic model database 130 for use by the phone recognizer 110 during the user's next utterance or interaction. The database 130 may be any mechanism, memory, file etc. capable of storing information and data known to those of skill in the art.

[0045] As shown in Figs. 1 and 2, the method of this invention may be implemented using a programmed processor. However, method can also be implemented on a general-purpose or a special purpose computer, a

5

[0046]

10

What is claimed is:

- 1 1. A method for building a phonotactic model for domain independent
2 speech recognition, comprising:
3 recognizing phones from a user's input communication using a
4 current phonotactic model stored in a database;
5 detecting morphemes from the recognized phones;
6 creating a new phonotactic model using the detected morphemes;
7 replacing the current phonotactic model with the new phonotactic
8 model in the database; and
9 outputting the detected morphemes for processing.
- 1 2. The method of claim 1, wherein the morphemes include at least
2 one of verbal speech and non-verbal speech.
- 1 3. The method of claim 2, wherein the non-verbal speech includes the
2 use of at least one of gestures, body movements, head movements, non-
3 responses, text, keyboard entries, keypad entries, mouse clicks, DTMF codes,
4 pointers, stylus, cable set-top box entries, graphical user interface entries and
5 touchscreen entries.
- 1 4. The method of claim 1, wherein the morphemes are expressed in
2 multimodal form.
- 1 5. The method of claim 1, wherein the user's input communication is
2 derived from the verbal and non-verbal speech and the user's environment.
- 1 6. The method of claim 1, wherein the morphemes in the user's input
2 communication are derived from the user's actions, including the user's focus of
3 attention.
- 1 7. The method of claim 1, further comprising:
2 mapping the detected acoustic morphemes into words.
- 1 8. The method of claim 1, wherein the method operates in conjunction
2 with one or more communication networks, the communication networks
3 including a telephone network, the Internet, an intranet, cable TV network, a local
4 area network (LAN), and a wireless communication network.

1 9. The method of claim 1, wherein the method is used for customer
2 care purposes.

1 10. The method of claim 1, wherein the morphemes are detected using
2 a measure of commonality within a language of the morphemes.

1 11. The method of claim 10, wherein the commonality measure is a
2 mutual information measure.

1 12. The method of claim 11, wherein each of the plurality of detected
2 morphemes has a mutual information measure exceeding a predetermined
3 threshold.

1 13. The method of claim 1, wherein the step of detecting acoustic
2 morphemes includes using a confidence function.

1 14. The method of claim 1, wherein the new phonotactic model is used
2 in the recognize step for a future input communication from a user.

1 15. The method of claim 1, wherein the method is iterative.

1 16. The method of claim 1, wherein the morphemes are at least one of
2 acoustic morphemes and non-acoustic morphemes.

1 17. A method of task classification using a phonotactic model built for
2 domain independent speech recognition, comprising:

3 recognizing phones from a user's input communication using a
4 current phonotactic model stored in a database;

5 detecting morphemes from the recognized phones;

6 creating a new phonotactic model using the detected morphemes;

7 replacing the current phonotactic model with the new phonotactic
8 model in the database; and

9 making task-type classification decisions based on the detected
10 morphemes from the user's input communication.

1 18. The method of claim 17, wherein the morphemes include at least
2 one of verbal speech and non-verbal speech.

1 19. The method of claim 18, wherein the non-verbal speech includes
2 the use of at least one of gestures, body movements, head movements, non-
3 responses, text, keyboard entries, keypad entries, mouse clicks, DTMF codes,
4 pointers, stylus, cable set-top box entries, graphical user interface entries and
5 touchscreen entries.

1 20. The method of claim 17, wherein the morphemes are expressed in
2 multimodal form.

1 21. The method of claim 17, wherein the user's input communication is
2 derived from the verbal and non-verbal speech and the user's environment.

1 22. The method of claim 17, wherein the morphemes in the user's input
2 communication are derived from the user's actions, including the user's focus of
3 attention.

1 23. The method of claim 17, further comprising entering into a dialog
2 with the user to obtain a feedback response from the user.

1 24. The method of claim 23, wherein the user is prompted to provide a
2 feedback response includes additional information with respect to the user's
3 initial input communication.

1 25. The method of claim 23, wherein the user is prompted to provide a
2 feedback response that includes confirmation with respect to at least one of the
3 set of task objectives determined in the classification decision.

1 26. The method of claim 17, wherein the input communication is routed
2 based on the classification decision.

1 27. The method of claim 26, wherein the task objective is performed
2 after the input communication is routed.

1 28. The method of claim 17, wherein the method operates in
2 conjunction with one or more communication networks, the communication
3 networks including a telephone network, the Internet, an intranet, cable TV
4 network, a local area network (LAN), and a wireless communication network.

1 29. The method of claim 17, wherein the method is used for customer
2 care purposes.

1 30. The method of claim 17, wherein the classification decisions and
2 corresponding user input communications are collected for automated learning
3 purposes.

1 31. The method of claim 17, wherein the relationship between the
2 detected morphemes and the predetermined set of task objectives includes a
3 measure of usefulness of a one of the morphemes to a specified one of the
4 predetermined task objectives.

1 32. The method of claim 17, wherein the relationship between the
2 detected morphemes and the predetermined set of task objectives includes a
3 measure of commonality within a language of the morphemes.

1 33. The method of claim 32, wherein the commonality measure is a
2 mutual information measure.

1 34. The method of claim 33, wherein each of the plurality of detected
2 morphemes has a mutual information measure exceeding a predetermined
3 threshold.

1 35. The method of claim 17, wherein the step of making a classification
2 decision includes a confidence function.

1 36. The method of claim 17, wherein the input communication from the
2 user represents a request for at least one of the set of predetermined task
3 objectives.

1 37. The method of claim 17, wherein the input communication is
2 responsive to a query of a of form "How may I help you?".

1 38. The method of claim 17, wherein each of the verbal and non-verbal
2 speech are directed to one of the set of predetermined task objectives and each
3 of the verbal and non-verbal speech is labeled with the one task objective to
4 which it is directed.

1 39. The method of claim 17, wherein the new phonotactic model is
2 used in the recognize step for a future input communication from a user.

1 40. The method of claim 17, wherein the method is iterative.

1 41. The method of claim 17, wherein the morphemes are at least one
2 of acoustic morphemes and non-acoustic morphemes.

1 42. A system that builds a phonotactic model for domain independent
2 speech recognition, comprising:

- 3 a database that stores phonotactic models;
- 4 a recognizer that recognizes phones from a user's input
- 5 communication using a current phonotactic model stored in the database;
- 6 a morpheme detector that detects morphemes from the recognized
- 7 phones and outputs the detected morphemes for processing; and
- 8 a learning unit that creates a new phonotactic model using the
- 9 detected morphemes, wherein the learning unit replaces the current phonotactic
- 10 model with the new phonotactic model in the database.

 43. The system of claim 42, further comprising:
 a task classification processor that makes task classification
decisions based on the detected morphemes from the user's input
communication.

5

Abstract of the Disclosure

The invention concerns a method and corresponding system for building a phonotactic model for domain independent speech recognition. The method may include recognizing phones from a user's input communication using a current
 10 phonotactic model, detecting morphemes (acoustic and/or non-acoustic) from the recognized phones, and outputting the detected morphemes for processing. The method also updates the phonotactic model with the detected morphemes and stores the new model in a database for use by the system during the next user interaction. The method may also include making task-type classification
 15 decisions based on the detected morphemes from the user's input communication.

107260 20095660