

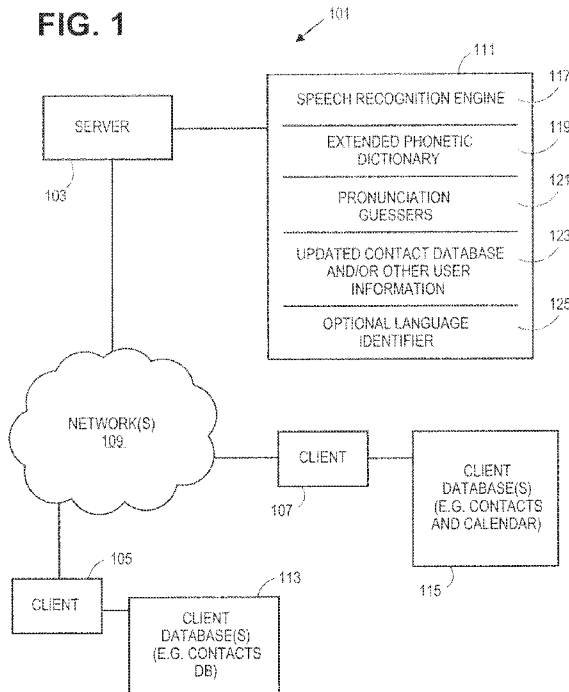


- (51) International Patent Classification:  
G10L 15/187 (2013.01) G10L 15/06 (2013.01)
- (21) International Application Number:  
PCT/US2013/044574
- (22) International Filing Date:  
6 June 2013 (06.06.2013)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
13/492,720 8 June 2012 (08.06.2012) US
- (71) Applicant: APPLE INC. [US/US]; 1 Infinite Loop, Cupertino, CA 95014 (US).
- (72) Inventor: NAIK, Devang, K.; 1 Infinite Loop, Cupertino, CA 95014 (US).
- (74) Agents: BREGMAN, Dion, M. et al.; Morgan Lewis & Bockius LLP, 2 Palo Alto Square, 3000 El Camino Real, Suite 700, Palo Alto, CA 94306 (US).

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

[Continued on next page]

(54) Title: SPOKEN NAMES RECOGNITION



(57) Abstract: A speech recognition system uses, in one embodiment, an extended phonetic dictionary that is obtained by processing words in a first user's set of databases, such as a first user's contacts database, with a set of pronunciation guessers. The speech recognition system can use a conventional phonetic dictionary and the extended phonetic dictionary to recognize speech inputs that are user requests to use the contacts database, for example, to make a phone call, etc. The extended phonetic dictionary can be updated in response to changes in the contacts database, and the set of pronunciation guessers can include pronunciation guessers for a plurality of locales, each locale having its own pronunciation guesser.

WO 2013/184953 A1



---

**Published:**

— with international search report (Art. 21(3))

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

## SPOKEN NAMES REGOGNITION

### BACKGROUND OF THE INVENTION

[0001] Embodiments of the present invention relate to the field of speech recognition. Speech recognition systems have been deployed for many years on various types of devices including desktop and laptop computer systems as well as telephone systems, such as cellular telephones and/or smartphones which include cellular telephones. One use of speech recognition systems in telephones, such as smart phones, is the use of name dialing which allows a user to speak a name in a contacts database in order to cause the telephone to initiate a telephone call. Speech recognition systems can use phonetic dictionaries or lexicons in order to recognize spoken words. Some speech recognition systems, particularly those which are adapted to provide natural language recognition, use a large phonetic dictionary to model a large set of words. These systems can be used to recognize names in an address book or contacts database, but these systems often have difficulty with names due to the fact that the names are often not modeled in a phonetic dictionary or the names are in a different locale from the locale used in the phonetic dictionary.

### SUMMARY OF THE DESCRIPTION

[0002] A speech recognition system can, in one embodiment of the invention, use an extended phonetic dictionary that is obtained by processing words in a user's set of databases, such as a user's contacts database or calendar or media database, etc., with a set of one or more pronunciation guessers. The speech recognition system can use a conventional phonetic dictionary and the extended phonetic dictionary to recognize speech inputs such as a speech input in which a user requests to use the contacts database, for example, to make a telephone call, etc. The extended phonetic dictionary can be updated in response to changes in the contacts database, and the set of pronunciation guessers can include pronunciation guessers for a plurality of locales, each locale having its own pronunciation guesser. In one embodiment, the extended phonetic dictionary can be updated at runtime based on changes to the user's databases, such as changes to the contacts databases, etc.

[0003] In one embodiment of the invention, a data processing system can perform a method which includes storing a phonetic dictionary for speech recognition; obtaining words from a user's set of one or more databases; processing, using a set of one or more

pronunciation guessers, the words from the user's set of one or more databases, the processing producing additional phonetic data derived from the words, the additional phonetic data forming an extended phonetic dictionary for the user; receiving a speech input; and processing the speech input by comparing phonemes detected in the speech input to the phonetic dictionary and to the extended phonetic dictionary to determine a set of one or more matches, from which a best match can be determined. Phonetic fuzzy matching can be used to determine the best match. In one embodiment, the user's set of one or more databases can include a contacts database, which can also be referred to as an address book, which includes names of people and telephone numbers and email addresses, etc. In one embodiment, the phonetic dictionary can be for a natural language speech recognition system, and the speech input from the user can be a request to call a telephone number or to send a text message to another user's device or to play a song or movie on the user's device. The user's set of one or more databases can also include one or more calendar databases, and databases for medias, such as songs or movies or pictures, etc. In one embodiment, the method can also include obtaining changes in the contacts database and processing, using the set of one or more pronunciation guessers, the changes to update the extended phonetic dictionary based on the changes; the obtaining of the changes can occur in response to the changes being made or can occur in response to a speech input which causes the system to determine whether or not such changes exist and to thereby then process those changes using the set of one or more pronunciation guessers. In one embodiment, the set of one or more pronunciation guessers include pronunciation guessers for a plurality of locales, each locale having its own pronunciation guesser. For example, if the user's device has been set for a USA English locale, then the pronunciation guessers can include, in one embodiment, pronunciation guessers for American English, Spanish, and potentially other languages which are commonly used in the United States of America (USA). If the user's device has been set for a locale in Europe, then the set of pronunciation guessers for a set of locales can include pronunciation guessers for English, French, German, and Italian. In one embodiment, the set of pronunciation guessers can include different types of pronunciation guessers for the same locale in order to provide a more robust selection of pronunciation guesses. In one embodiment, a locale can be considered to be a language or a dialect of a language.

**[0004]** In one embodiment, the method can also include determining automatically a set of languages or locales from the words in the user's set of one or more databases. Automatic language detection systems and modules currently exist which can analyze words in a document or database and determine, from that analysis, the languages used. An

example of such an automatic language identification system is provided in published U.S. Application US2009/0307584. A subset of the set of pronunciation guessers can be selected based upon the languages identified by an automatic language identifier system, and this selected subset can then be used to process words in the user's databases, such as contacts databases, calendar databases, media databases, etc., in order to generate an extended phonetic dictionary based upon a selected subset of locales which are derived from the analysis performed by the automatic language identifier. In one embodiment, the processing of a speech input can use only that subset of pronunciation identifiers which are selected as a result of the analysis performed by the automatic language identifier.

**[0005]** In one embodiment, the method can be performed by a server that is coupled through a wireless network to the user's device which includes the contacts database and other databases in the user's set of databases. The wireless network can be a local WiFi network or a cellular telephone network or a combination of such wireless networks. The server can obtain the words in the user's databases through the wireless network and can receive the speech input from the user's device through the wireless network, via a network interface (at the server) which is coupled to the wireless network. The server can transmit the best match as the result of the speech recognition operation to the user's device through the wireless network in this embodiment.

**[0006]** In another embodiment, all of the operations in a method according to an embodiment of the present invention can be performed on the user's device without the use of a server. In yet another embodiment, the user's device and the server can split the processing between the two devices and perform various parts of the methods described herein on each of the two devices.

**[0007]** In accordance with some embodiments, an electronic device includes an input device for receiving speech input, one or more processors, memory, and one or more programs; the one or more programs are stored in the memory and configured to be executed by the one or more processors and the one or more programs include instructions for performing the operations of any of the methods described above. In accordance with some implementations, a computer readable storage medium has stored therein instructions which when executed by an electronic device with an input device for receiving speech input, a touch-sensitive surface, and optionally one or more sensors to detect intensity of contacts with the touch-sensitive surface, cause the device to perform the operations of any of the methods referred described above. In accordance with some implementations, an electronic

device includes: an input device for receiving speech inputs; and means for performing the operations of any of the methods described above. In accordance with some implementations, an information processing apparatus, for use in an electronic device with an input device for receiving speech input, includes means for performing the operations of any of the methods described above.

**[0008]** In accordance with some embodiments, an electronic device includes a speech receiving unit configured to receive a speech input; a user database unit configured to store a first user's set of one or more databases; a phonetic dictionary storage unit configured to store a phonetic dictionary for speech recognition; and a processing unit coupled to the speech receiving unit and the phonetic dictionary storage unit. The processing unit is configured to obtain words from the first user's set of one or more databases; process, using a set of one or more pronunciation guessers, the words from the first user's set of one or more databases, the processing producing additional phonetic data derived from the words, the additional phonetic data forming an extended phonetic dictionary for the first user; acquire a speech input from the speech receiving unit; process the speech input by comparing phonemes detected in the speech input to the phonetic dictionary and to the extended phonetic dictionary to determine a set of one or more matches; and determine a best match from the set of one or more matches.

**[0009]** The embodiments described herein can be implemented as machine readable non-transitory storage media or as methods or as one or more data processing systems.

**[0010]** The above summary does not include an exhaustive list of all aspects of the present invention. It is contemplated that the invention includes all systems and methods that can be practiced from all suitable combinations of the various aspects summarized above, and also those disclosed in the Detailed Description below.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

**[0011]** The present invention is illustrated by way of example and not limitation in the figures of the accompanying drawings in which like references indicate similar elements.

**[0012]** Figure 1 shows an example of a system in which a server provides speech recognition services to a plurality of client devices according to one embodiment of the present invention.

[0013] Figure 2 shows an example of an embodiment of the invention in which a client data processing system implements all aspects or a portion of the aspects of one or more embodiments described herein.

[0014] Figure 3 provides an example of the use of two or more pronunciation guessers according to one embodiment of the present invention.

[0015] Figure 4 provides an example of the use of two pronunciation guessers for two different locales according to another embodiment of the present invention.

[0016] Figure 5 is a flowchart which illustrates a method according to one embodiment of the present invention.

[0017] Figure 6 is a flowchart which illustrates another method according to an embodiment of the present invention.

[0018] Figure 7 is a flowchart which illustrates a method according to an embodiment of the present invention.

[0019] Figure 8 is a flowchart which shows a method according to one embodiment of the present invention.

[0020] Figure 9 is a block diagram of a data processing system which can be used to implement one or more embodiments of the present invention.

[0021] Figure 10 is a functional block diagram of an electronic device in accordance with some embodiments.

## DETAILED DESCRIPTION

[0022] Various embodiments and aspects of the inventions will be described with reference to details discussed below, and the accompanying drawings will illustrate the various embodiments. The following description and drawings are illustrative of the invention and are not to be construed as limiting the invention. Numerous specific details are described to provide a thorough understanding of various embodiments of the present invention. However, in certain instances, well-known or conventional details are not described in order to provide a concise discussion of embodiments of the present inventions.

[0023] Reference in the specification to “one embodiment” or “an embodiment” means that a particular feature, structure, or characteristic described in conjunction with the

embodiment can be included in at least one embodiment of the invention. The appearances of the phrase “in one embodiment” in various places in the specification do not necessarily all refer to the same embodiment. The processes depicted in the figures that follow are performed by processing logic that comprises hardware (e.g. circuitry, dedicated logic, etc.), software, or a combination of both. Although the processes are described below in terms of some sequential operations, it should be appreciated that some of the operations described may be performed in a different order. Moreover, some operations may be performed in parallel rather than sequentially.

**[0024]** Figure 1 shows a client/server system in which a server system provides speech recognition services for one or more client devices, each of which can be a data processing system such as a smartphone or other cellular telephones. In the client/server system 101 shown in Figure 1, at least two client devices 105 and 107 are shown coupled through one or more networks 109 to a server system 103. Server system 103 can be a general purpose computer system or a special purpose computer system or other types of data processing systems (such as the system shown in Figure 9) which can be configured to provide speech recognition services; for example these systems can be configured by using software that programs the data processing system to perform the speech recognition services. This software can include a speech recognition engine 117 stored in a storage device 111. The speech recognition engine 117 can, in one embodiment, be a natural language speech recognition system that uses a conventional phonetic lexicon such as a phonetic dictionary as is known in the art. Storage device 111 also includes other components which can be used by the server system 103 and which will be described further below. In the system shown in Figure 1, each of the clients, such as clients 105 and 107 can send a digitized speech input to the server system 103 which then processes the digitized speech input using the speech recognition system at the server system 103. Each of the clients, such as clients 105 and 107 can be a desktop computer system or a laptop computer system or a cellular telephone, or a smartphone, or other data processing systems or consumer electronic devices, including embedded electronic devices. These clients can include one or more databases for each of the clients. For example, client 105 can be a smartphone for a particular user that includes databases for that user such as a contacts or address book database and a calendar database, and media databases, and potentially other databases storing information for that particular user, which can be unique for that particular user relative to other users of other client systems.



**[0025]** In the example shown in Figure 1, the client databases 113 can be stored on storage device, such as flash memory of client 105 and can be local storage only at client 105 system; in another embodiment, the client databases, such as client database 113 can be stored remotely on a server system which is accessible through a network. Similarly, client 107 includes client databases 115 which can be unique for the user of client system 107 and can include a contacts database and a calendar database and potentially other databases. Each of the client systems, such as client systems 105 and 107 can include a speech input hardware system such as one or more microphones designed to capture the user's speech input and then a system which processes and digitizes that speech input and then transmits that speech input through one or more networks 109 to the server system 103. In one embodiment, one or more of the client systems, such as client systems 105 or 107 or a server system such as server 103 may include a speech repair system, such as the system described in pending U.S. Patent Application Serial No. 13/247,912, filed September 28, 2011, and entitled "Speech Recognition Repair Using Contextual Information", and this application is hereby incorporated herein by reference.

**[0026]** Each of the clients in system 101 can be configured to transmit data representing the content of each of the client databases so that the server system 103 can process the words in the databases according to the methods described herein using the pronunciation guessers to generate one or more extended phonetic dictionaries according to the embodiments described herein.

**[0027]** The one or more networks 109 can be one or more cellular telephone networks or one or more local area wireless networks, such as one or more WiFi networks or a combination of such networks and can include both wired and wireless portions. The one or more networks 109 serve to provide a data connection to transmit data, such as words in the client databases, such as names and telephone numbers in the contacts databases of the various users using the client devices, and the one or more networks 105 can also provide, in one embodiment, voice telephony services allowing users to talk over their cellular telephones as well as provide text messaging services and other services known in the art in connection with cellular telephones or smartphones.

**[0028]** Storage device 111 is coupled to server system 103 either as part of the server system 103 (e.g., a local hard drive of server system 103) or as a storage device on a network which is coupled to server system 103. Server system 103 can be a set of servers in a server farm and storage device 111 can be a set of storage devices in a Storage Area Network

(SAN). Storage device 111 can be one or more hard drives containing software and data to provide the speech recognition services described in conjunction with the one or more embodiments herein. Speech recognition engine, as noted above, can be a conventional speech recognition system which uses a conventional phonetic dictionary. In addition, the present invention provides an extension to that phonetic dictionary through one or more extended phonetic dictionaries 119. In one embodiment, each user which is provided speech recognition services by server system 103 can have their own dedicated extended phonetic dictionary 119 which is produced according to one or more methods described herein. The server system 103 can also include one or more pronunciation guessers 121 which are used to process words in the client databases retrieved from (e.g., dynamically uploaded from) client devices. As shown in Figure 1, storage device 111 can include an updated contacts database and/or other user information for each of the users of client devices, such as client devices 105 and 107. In one embodiment, each client device can upload the latest version of contact databases and other databases from each client device or can upload just the last set of changes that were made since the last upload of the client database or changes to the client database. In one embodiment, each client device can upload a contacts database, a calendar database, a media database, and potentially other databases containing unique user information, and these databases are stored as, for example, updated contact database 123 in the storage device 111. In one embodiment, storage device 111 can also include one or more optional language identifier systems, such as optional language identifier 125 which can be a software module that performs automatic language identification, such as the automatic language identification described in published U.S. Patent Application US2009/0307584 which application is hereby incorporated herein by reference. The optional language identifier can analyze the words in the client's databases to determine the languages or locales used in the those databases and to thereby select the subset of pronunciation guessers which are appropriate for those locales and which match those locales; further aspects with respect to automatic language identification are described further below.

**[0029]** While Figure 1 shows an example of a client/server architecture in which a server performs at least a portion of the speech recognition services according to one embodiment, Figure 2 represents an alternative embodiment in which a client device performs all those services without the need for a server system to perform speech recognition services. In this embodiment, a client data processing system such as the system 201 can have all the resources stored on a local storage device 201. These resources can include a speech recognition engine 203 which can be, in one embodiment, software and data

designed to provide natural language speech recognition and can include a standard or conventional phonetic dictionary used in the process of providing natural language speech recognition. In addition, storage device 201 can include an extended phonetic dictionary 205 which is derived from one or more user databases 209, such as a contacts database or a calendar database or a media database or other databases specific to and unique to the user of the client system 201. The extended phonetic dictionary 205 can be derived by using one or more pronunciation guessers, such as pronunciation guessers 207 which process words in the one or more client databases 209 on the storage device 201. In one embodiment, the client databases can include a contacts or address book database or other data structures that include names, telephone numbers, addresses, email addresses, and other contact information and other pertinent information for one or more people or businesses for the user of the data processing system 201. In addition, other databases can include a calendar, database or a plurality of calendar databases, a media database, such as a song, movie, etc. database, and other user databases which may be the subject of speech input commands from the user of data processing system 201. Data storage 202 may also include an optional language identifier 211 which can be an automatic language identification module, such as the systems described in published U.S. Patent Application 2009/0307584.

**[0030]** It will be appreciated that the system shown in Figures 1 or 2 can be employed with a variety of the different embodiments described herein, such as the embodiments described in conjunction with Figures 5, 6, 7, and 8 as well as other embodiments described herein. Moreover it will be appreciated that the processing operations in any one of the methods described herein may be split among a client and server architecture such that a client system can perform some of the operations while the server performs other of the operations in order to achieve a result of any one of the methods described herein.

**[0031]** Figures 3 and 4 show two examples of the use of a plurality of pronunciation guessers when words in a user's database, such as a contacts or address book database are being processed. In the example shown in Figure 3, an input word "Devang" is used as an input to two or more pronunciation guessers. This input word 303 results in three different outputs from three different pronunciation guessers 305 and 307 and 309. In particular, outputs 306 and 308 and 310 result from the processing performed by three different pronunciation guessers, guessers 305 and 307 and 309 respectively, based upon the input. The three different pronunciation guessers can be pronunciation guessers that use different techniques in processing the inputs to derive an output. There are a plurality of known

pronunciation guessers that can be used in various embodiments of the present invention. For example, one known pronunciation guesser uses a classification and regression tree to guess the pronunciation of words which in effect converts letters to sound rules. Another example of a known pronunciation guesser which can convert letters to sound rules is a pronunciation guesser which employs a maximum likelihood estimation technique. Yet another pronunciation guesser which can be used is a character language modeling pronunciation guesser which can also provide an output at a phoneme level based upon an input word. In the example shown in Figure 3, each of the pronunciation guesser 305 and 307 and 309 may be for the same locale, such as an American English locale or a French locale or a European Spanish locale. In another embodiment, the example of Figure 3 shows the use of pronunciation guessers for three different locales (for example, A = USA locale; B = Spanish locale; and C = French locale), and each of these pronunciation guessers can use the same technique (for example, maximum likelihood estimation) for their corresponding locale.

**[0032]** Figure 4 shows another example which can be employed in various embodiments of the present invention in which an input word from a user's database, such as a contacts database, is processed using two different pronunciation guessers for two different locales. In this case, the same pronunciation guesser technique, such as a maximum likelihood estimation technique can be used for two different locales. This approach is advantageous when a user's device is likely to contain names or other words that are from different locales. For example, it is often the case in the United States that individual names in a user's contacts or address book can be Spanish names or Chinese names or certain other locales. In this case, it is advantageous that pronunciation guessers for different locales be employed to analyze the names and words in the user's databases in order to generate different outputs which can then be matched using known techniques, such as phonetic fuzzy matching techniques, in order to derive a best match from a set of possible matches, as will be explained further below. In this example shown in Figure 4, an input word 403 is provided as an input to two pronunciation guessers for two different locales. For example, one locale can be for American English and another locale can be for American Spanish. Each guesser for each locale can then provide an output, such as the outputs 406 and 408 from the guessers 405 and 407 respectively as shown in Figure 4 and these outputs can be mapped, as explained herein, to the target locale of the standard phonetic dictionary. These outputs can then be used as part of the extended phonetic dictionary for the user when speech recognition is performed. In one embodiment, these extended phonetic dictionaries can be embedded within a conventional or standard phonetic dictionary or can be maintained and

processed separately from a standard or conventional phonetic dictionary. In either case, the use of the extended phonetic dictionary can improve the accuracy of speech recognition when the speech recognizer is attempting to recognize names in one or more locales in a user's set of databases, such as a contacts database. In one embodiment, a standard phonetic dictionary will be configured or designed to work in a particular locale which can be referred to as a target locale; for example, a user who lives in the USA and speaks American English would have a system that uses a standard phonetic dictionary that uses American English as the target locale. Further, pronunciation guessers for other locales that were used to create the extended phonetic dictionary may use a mapping from phonemes in the other locales to phonemes in the target locale, and this mapping can be done using an aligner or a set of manually setup rules that were prepared by a linguist who is familiar with the languages used in the target locale and one or more of the other locales. For example, if a Finnish pronunciation guesser is used to create outputs for the extended phonetic dictionary then a mapping from Finnish phonemes to phonemes in the target locale can be used to create the final output for the extended phonetic dictionary.

**[0033]** Figures 5 and 6 illustrate two exemplary methods for maintaining an extended phonetic dictionary even after the user makes one or more changes to the user's databases. For example, the methods shown in Figures 5 and 6 can maintain the extended phonetic dictionary while the user adds names to a contacts database or changes the names (for example when someone gets married) or deletes names from their contacts or address book. A method as in Figure 5 can begin in operation 501 in which the data processing system obtains a user's one or more databases such as the user's contacts database and calendar database, and/or media database. These databases can be obtained by uploading the databases from the user's device through a network to a server, such as the server system 103 in the case of Figure 1 or can be obtained by localizing retrieving the databases from the user's data processing system as in the case of Figure 2. It will be appreciated that the term "database" is meant to cover any data structure containing the user data such as contacts or address book data or calendar data or media data or other types of user data. These data structures can be a structured data structure in which data is entered into fields or can be an unstructured data structure as in, for example, a sequence or string of text. Then in operation 503, the data processing system can use at least one pronunciation guesser, such as a default guesser for the currently selected locale on the data processing system, to process words in databases to produce an extended phonetic dictionary for the user. Figure 4 shows one such example in which a set of pronunciation guessers for different locales is used to process

words in the user's databases. In the case of the architecture shown in Figure 1, the server system 103 can perform operation 503 by processing words in the databases, after the databases have been retrieved through the uploading process. In the case of the architecture shown in Figure 2, the client data processing system can perform operation 503 by using its local versions of one or more pronunciation guessers 207 to create a local extended phonetic dictionary 205 which is derived from the processing of the pronunciation guesser 207 on words in client database 209. After operation 503, the system can then store the extended phonetic dictionary for the user in operation 505; the storage of the extended phonetic dictionary can either be a local storage as in the case of the architecture shown in Figure 2 or a remote storage, relative to the user's device, as in the case of the architecture shown in Figure 1. After an initial or updated version of an extended phonetic dictionary has been stored, the system in operation 507 can determine whether or not the user's databases have changed. Operation 507 may be performed as a periodic or intermittent task (e.g. at every system start-up or wake from sleep) or may be performed in response to a notification from a daemon software component that a user's database has changed or can be performed in response to a speech input which can then determine whether or not the user's database has changed since the last time that the extended phonetic dictionary was updated. If no change has occurred then, the extended phonetic dictionary last created can be maintained in storage, but if a change has occurred in one of the user's databases, then operation 509 can be performed in which the extended phonetic dictionary is updated by performing operation 503 again and storing the result of the updated extended phonetic dictionary. Operation 509 can involve using at least one pronunciation guesser on the entire user's databases without regard the changes (if the prior extended phonetic dictionary is to be discarded) or can be performed on just the portions of the databases that have changed.

**[0034]** Figure 6 shows an example of another method for creating and maintaining an extended phonetic dictionary for a user's one or more databases, such as contacts or calendars or media databases or a combination of such databases. In operation 601, a data processing system can obtain the user's one or more databases from either a local storage as in the case of the architecture shown in Figure 2 or by receiving the databases when the client device uploads the one or more databases as in the architecture shown in Figure 1. After the databases are obtained, operation 603 can optionally be performed in order to determine the locales that are used in the databases. This can be performed with the use of automatic language identifiers which are known in the art; see for example published U.S. Patent Application US2009/0307584. These automatic language identifiers can process the words

in the user's one or more databases and determine the locales based on those words. The result of operation 603, if performed, can allow a system to select a subset of available pronunciation guessers based on the identified languages so that the system need only use that subset rather than all of the pronunciation guessers which are available to the system. For example, if the system has ten pronunciation guessers for ten different locales, and the language identifier determines that only three locales are used in the user's databases, then only those three pronunciation guessers can be used when processing the words in the user's database in order to create the extended phonetic dictionary. If operation 603 is not performed, then the system can use a default set of pronunciation guessers based upon a currently selected locale. For example, if the United States of America is the currently selected locale, then the default pronunciation guessers could be a pronunciation guesser for American English and American Spanish (i.e., Spanish spoken in the USA), and no other pronunciation guessers are used when processing the user's set of one or more databases to create the extended phonetic dictionary. In operation 605, the data processing system processes the words in the databases to produce the extended phonetic dictionary for the user, again using either a default set of pronunciation guessers or a selected subset of pronunciation guessers. It will be appreciated that in addition to using different pronunciation guessers for different locales, the system can also use different types of pronunciation guessers. As described herein there are at least three different types of pronunciation guessers which are currently available. After the words are processed in operation 605, the system can store the extended phonetic dictionary for the user in operation 607, and then operation 609, which resembles operation 507, can be performed. In response to determining in operation 609 that the user's database has changed, then, in operation 611, the extended phonetic dictionary is updated based upon those changes by either processing just the changes or by repeating the entire processing of the entire database and discarding the previously created extended phonetic dictionary.

**[0035]** Figures 7 and 8 provide examples of methods in which the extended phonetic dictionary is used in conjunction with speech recognition. In one embodiment, the speech recognition system used in both Figures 7 and 8 can be a natural language speech recognition system which uses a large and conventional phonetic dictionary as part of the speech recognition engine, such as speech recognition engine 117 in Figure 1 or 203 in Figure 2. The method shown in Figure 7 can begin in operation 701 in which a data processing system receives a speech input from the user; for example the user can say (by speaking into a microphone of a client device) "call Devang" and the client system can receive and digitize

the speech input using known techniques and then process the speech input using a speech recognition engine such as the speech recognition engine 117 or the speech recognition engine 203. In the case of the use of speech recognition 203, the speech input is received and processed at the client data processing system 201, whereas in the embodiment shown in Figure 1, the speech input is received and digitized and then transmitted through network 109 to the server system 103 which then processes the digitized speech input, after having received it through a network interface at the server system 103. The speech recognition engine can then generate in operation 703 an initial set of recognized phonemes from the speech input received in operation 701. Then the system can in operation 705 use the standard phonetic dictionary and the extended phonetic dictionary (which can be embedded in the standard phonetic dictionary) for the user to determine possible matches. In one embodiment, the determining of possible matches can use a technique to obtain a priori knowledge about where, in a recognized sequence of phonemes, names are expected to be and to use the extended phonetic dictionary on at least the portion of phonemes where names are expected to be. For example, as explained in the “Speech Recognition Repair Using Contextual Information” application (No. 13/247,912, filed September 28, 2011) referred to herein, contextual information can be used to predict the location of a name between two recognized words “call” and “mobile” when the user says “call Devang mobile”. In this case, the recognized words “call” and “mobile” can be used to determine that the phonemes between those recognized words are likely to be phonemes for a name; in other words, the system can use the context of the words “call” and “mobile” (as indicating a request to make a phone call to someone’s mobile phone) to infer that the phonemes between those words are for a name and then use the extended phonetic dictionary, having multiple locales, on at least those phonemes. In operation 707, the system can then select the best match as a final recognized output; the selection of a best match can utilize phonetic fuzzy matching techniques which are known in the art. Then in operation 709 a client data processing system, such as client system 105 or 107 or 201 can optionally display the recognized output from operation 707 to the user to allow the user to confirm or to retry by speaking the input again or to cancel the operation requested by the user.

**[0036]** Figure 8 shows another example of a method according to an embodiment of the invention in which speech is processed using an extended phonetic dictionary in combination with a standard phonetic dictionary. In operation 801, the speech input is received from the user; this input can be, for example, “call Devang at home”. The words “call” and “at” and “home” can be recognized by the standard phonetic dictionary and the



word “Devang” can be recognized by the extended phonetic dictionary. Operation 801 can be similar to operation 701 in which the speech input can be received at a client device and then transmitted through a network to a server system or can be received at the client device and processed at the client device as in the architecture shown in Figure 2. In response to operation 801, the system can then in operation 803 process any changes in the user’s databases and update the extended phonetic dictionary based on those changes. In the embodiment shown in Figure 8, it is assumed that the data processing system will be fast enough to perform operation 803 after having received the speech input. Then in operation 805 the system generates an initial set of recognized phonemes from the received speech input. In operation 807, these recognized phonemes are then processed using a standard phonetic dictionary and the updated extended phonetic dictionary for the user to determine possible matches. The best match can then be selected in operation 809 by using, for example, conventional phonetic fuzzy matching to derive the final output from the speech recognition system. As in the method shown in Figure 7, the method of Figure 8 can also include an optional operation in which the recognized output is displayed to the user to allow the user to confirm the operation requested in the speech input or to retry the speech input process again or to cancel operation.

**[0037]** Figure 9 shows an example of data processing system 900 which may be used with one or more embodiments of the present invention. For example and in one embodiment, system 900 may be implemented as a portable data processing device such as a smartphone or tablet (e.g., iPad) device or a laptop or an entertainment system and client systems 105, 107, and 201 can be implemented as shown in Figure 9. The data processing system 900 shown in Figure 9 includes a processing system 911, which may be one or more microprocessors or which may be a system on a chip (integrated circuit) and the system also includes memory 901 for storing data and programs for execution by the processing system. The memory 901 can store, for example, the software components described in conjunction with Figures 1 or 2 and memory 901 can be any known form of a machine readable non-transitory storage medium, such as semiconductor memory (e.g., flash; DRAM; SRAM; etc.) The system 900 also includes an audio input/output subsystem 905 which may include a microphone and a speaker for, for example, playing back music or providing telephone functionality through the speaker and microphone. The microphone can receive the speech input described herein and that input can be digitized and provided to a speech recognizer system as described herein.

**[0038]** A display controller and display device 909 can provide a visual user interface for the user; this interface may include a graphical user interface which is similar to that shown on a Macintosh computer when running OS X operating system software or iOS software on an iPhone or iPad. The system 900 also includes one or more wireless transceivers 903 to communicate with another data processing system. A wireless transceiver may be a WLAN transceiver (e.g. WiFi), an infrared transceiver, a Bluetooth transceiver, and/or a wireless cellular telephony transceiver. It will be appreciated that additional components, not shown, may also be part of the system 900 in certain embodiments, and in certain embodiments fewer components than shown in Figure 9 may also be used in a data processing system. The system 900 further can include one or more communications ports 917 to communicate with another data processing system. The communications port may be a USB port, Firewire port, Bluetooth interface, a docking port, etc.

**[0039]** The data processing system 900 also includes one or more input devices 913 which are provided to allow a user to provide input to the system. These input devices may be a keypad or a keyboard or a touch panel or a multi-touch panel which is overlaid and integrated with a display device such as display device 909. The data processing system 900 can also include an optional input/output device which may be a connector for a dock. It will be appreciated that one or more buses, not shown, may be used to interconnect the various components as is well known in the art. The data processing system shown in Figure 9 may be a handheld computer or a personal digital assistant (PDA), or a cellular telephone with PDA-like functionality, or a handheld computer which includes a cellular telephone, or a media player, such as an iPod, or a game or entertainment device, or devices which combine aspects or functions of these devices, such as a media player combined with a PDA and a cellular telephone in one device or an embedded device or other consumer electronic devices. In other embodiments, the data processing system 900 may be a network computer or an embedded processing device within another device, or other types of data processing systems which have fewer components or perhaps more components than that shown in Figure 9.

**[0040]** Data processing system 900 can optionally include one or more hardware devices designed to digitize and store human speech received by the microphone in Audio I/O 905.

**[0041]** At least certain embodiments of the inventions may be part of a digital media player, such as a portable music and/or video media player, which may include a media

processing system to present the media, a storage device to store the media and may further include a radio frequency (RF) transceiver (e.g., an RF transceiver for a cellular telephone) coupled with an antenna system and the media processing system. In certain embodiments, media stored on a remote storage device may be transmitted to the media player through the RF transceiver. The media may be, for example, one or more of music or other audio, still pictures, or motion pictures.

**[0042]** Examples of a portable media player are described in published U.S. patent number 7,345,671 and U.S. published patent application number 2004/0224638, both of which are incorporated herein by reference.

**[0043]** In accordance with some embodiments, Figure 10 shows a functional block diagram of an electronic device 1000 configured in accordance with the principles of the invention as described above. The functional blocks of the device may be implemented by hardware, software, or a combination of hardware and software to carry out the principles of the invention. It is understood by persons of skill in the art that the functional blocks described in Figure 10 may be combined or separated into sub-blocks to implement the principles of the invention as described above. Therefore, the description herein may support any possible combination or separation or further definition of the functional blocks described herein.

**[0044]** As shown in Figure 10, the electronic device 1000 includes a speech receiving unit 1002 configured to receive a speech input. In some embodiments, the speech input includes a request that a telephone call be placed to another user's device. The electronic device 1000 also includes a user database unit 1004 configured to store a first user's set of one or more databases. The electronic device 1000 also includes a phonetic dictionary storage unit 1005 configured to store a phonetic dictionary for speech recognition. The electronic device 1000 also includes a processing unit 1006 coupled to the speech receiving unit 1002, the user database unit 1004, and the phonetic dictionary storage unit 1005. In some embodiments, the processing unit 1006 includes a word obtaining unit 1008, a word processing unit 1010, a speech processing unit 1012, and a best match determining unit 1014. In some embodiments, the word processing unit 1010 includes pronunciation guessers 1011.

**[0045]** The processing unit 1006 is configured to: obtain words from the first user's set of one or more databases (e.g., with the word obtaining unit 1008 in communication with the user database unit 1004). The processing unit 1006 is further configured to process, using a set of one or more pronunciation guessers (e.g., with the word processing unit 1010 and

pronunciation guessers 1011), the words from the first user's set of one or more databases, the processing producing additional phonetic data derived from the words, the additional phonetic data forming an extended phonetic dictionary for the first user. The processing unit 1006 is further configured to acquire a speech input from the speech receiving unit, and process the speech input (e.g., with the speech processing unit 1012) by comparing phonemes detected in the speech input to the phonetic dictionary and to the extended phonetic dictionary to determine a set of one or more matches. The processing unit 1006 is further configured to determine a best match from the set of one or more matches (e.g., with the best match determining unit 1014).

**[0046]** In some embodiments, the first user's set of one or more databases comprises a contacts database (e.g., contacts database 1007) with names of people and telephone numbers, and wherein the phonetic dictionary is for natural language speech recognition, and the speech input includes a request that a telephone call be placed to another user's device.

**[0047]** In some embodiments, the processing unit 1006 is further configured to: obtain changes in the contacts database made as a result of the production of the additional phonetic data (e.g., with the word processing unit 1010); and process, using the set of one or more pronunciation guessers (e.g., with the word processing unit 1010 and the pronunciation guessers 1011), the changes to update the extended phonetic dictionary based on the changes, wherein the obtaining of the changes occurs in response to the changes being made.

**[0048]** In some embodiments, the set of one or more pronunciation guessers (e.g., the pronunciation guessers 1011) comprise pronunciation guessers for a plurality of locales, each locale having its own pronunciation guesser. In some embodiments, the set of one or more pronunciation guessers (e.g., the pronunciation guessers 1011) comprises different pronunciation guessers for a same locale.

**[0049]** In some embodiments, the processing unit 1006 is further configured to: determine, automatically, a set of languages from the words in the first user's set of one or more databases (e.g., with the word processing unit 1010); and select a subset of the set of one or more pronunciation guessers (e.g., with the word processing unit 1010), the subset selected based upon the set of languages that were automatically determined from the words in the first user's set of one or more databases; wherein the processing which produces the additional phonetic data uses only the subset of the set of one or more pronunciation guessers when producing the additional phonetic data.

[0050] In some embodiments, the electronic device 1000 is the device of the first user which includes the contacts database (e.g., the contacts database 1007).

[0051] In the foregoing specification, the invention has been described with reference to specific exemplary embodiments thereof. It will be evident that various modifications may be made thereto without departing from the broader spirit and scope of the invention as set forth in the following claims. The specification and drawings are, accordingly, to be regarded in an illustrative sense rather than a restrictive sense.

What is claimed is:

1. A machine readable non-transitory storage medium storing executable instructions which, when executed by a data processing system, cause the data processing system to perform a method comprising:
  - storing a phonetic dictionary for speech recognition;
  - obtaining words from a first user's set of one or more databases;
  - processing, using a set of one or more pronunciation guessers, the words from the first user's set of one or more databases, the processing producing additional phonetic data derived from the words, the additional phonetic data forming an extended phonetic dictionary for the first user;
  - receiving a speech input;
  - processing the speech input by comparing phonemes detected in the speech input to the phonetic dictionary and to the extended phonetic dictionary to determine a set of one or more matches; and
  - determining a best match from the set of one or more matches.
2. The medium of claim 1 wherein the first user's set of one or more databases comprises a contacts database with names of people and telephone numbers, and wherein the phonetic dictionary is for natural language speech recognition, and wherein the speech input includes a request that a telephone call be placed to another user's device.
3. The medium of claim 2, wherein the method further comprises:
  - obtaining changes in the contacts database made as a result of the production of the additional phonetic data; and
  - processing, using the set of one or more pronunciation guessers, the changes to update the extended phonetic dictionary based on the changes, wherein the obtaining of the changes occurs in response to the changes being made.
4. The medium of any of claims 1-3, wherein the set of one or more pronunciation guessers comprise pronunciation guessers for a plurality of locales, each locale having its own pronunciation guesser.

5. The medium of any of claims 1-3, wherein the set of one or more pronunciation guessers comprises different pronunciation guessers for a same locale.
6. The medium of any of claims 1-5, wherein the method further comprises:
  - determining automatically a set of languages from the words in the first user's set of one or more databases; and
  - selecting a subset of the set of one or more pronunciation guessers, the subset selected based upon the set of languages that were automatically determined from the words in the first user's set of one or more databases;
  - wherein the processing which produces the additional phonetic data uses only the subset of the set of one or more pronunciation guessers when producing the additional phonetic data.
7. The medium of any of claims 2-6 wherein the method is performed by a server that is coupled through a wireless network to a device of the first user which includes the contacts database, and wherein the server obtains the words in the contacts database from the device of the first user through the wireless network, and wherein the server receives the speech input from the device of the first user through the wireless network, and wherein the server transmits the best match to the device of the first user through the wireless network.
8. The medium of any of claims 2-6 wherein the method is performed by the device of the first user which includes the contacts database.
9. The medium of claim 7 wherein the server performs the method for a plurality of users, and each user in the plurality of users has a dedicated extended phonetic dictionary that is unique to each user.
10. A machine implemented method comprising:
  - storing a phonetic dictionary for speech recognition;
  - obtaining words from a first user's set of one or more databases;
  - processing, using a set of one or more pronunciation guessers, the words from the first user's set of one or more databases, the processing producing additional phonetic data derived from the words, the additional phonetic data forming an extended phonetic dictionary for the first user;
  - receiving a speech input;

processing the speech input by comparing phonemes detected in the speech input to the phonetic dictionary and to the extended phonetic dictionary to determine a set of one or more matches; and

determining a best match from the set of one or more matches.

11. The method of claim 10 wherein the first user's set of one or more databases comprises a contacts database with names of people and telephone numbers, and wherein the phonetic dictionary is for natural language speech recognition and wherein the speech input includes a request that a telephone call be placed to another user's device.

12. The method of claim 11, wherein the method further comprises:

obtaining changes in the contacts database made as a result of the production of the additional phonetic data; and

processing, using the set of one or more pronunciation guessers, the changes to update the extended phonetic dictionary based on the changes, wherein the obtaining of the changes occurs in response to the changes being made.

13. The method of any of claims 10-12, wherein the set of one or more pronunciation guessers comprise pronunciation guessers for a plurality of locales, each locale having its own pronunciation guesser.

14. The method of any of claims 10-12, wherein the set of one or more pronunciation guessers comprises different pronunciation guessers for a same locale.

15. The method of any of claims 10-14, wherein the method further comprises:

determining automatically a set of languages from the words in the first user's set of one or more databases; and

selecting a subset of the set of one or more pronunciation guessers, the subset selected based upon the set of languages that were automatically determined from the words in the first user's set of one or more databases;

wherein the processing which produces the additional phonetic data uses only the subset of the set of one or more pronunciation guessers when producing the additional phonetic data.



16. The method of any of claims 11-15 wherein the method is performed by a server that is coupled through a wireless network to a device of the first user which includes the contacts database and wherein the server obtains the words in the contacts database from the device of the first user through the wireless network and wherein the server receives the speech input from the device of the first user through the wireless network and wherein the server transmits the best match to the device of the first user through the wireless network.

17. The method of any of claims 11-15 wherein the method is performed by the device of the first user which includes the contacts database.

18. An electronic device, comprising:

an input device for receiving speech input;

at least one processor; and

memory storing: a phonetic dictionary for speech recognition; and one or more programs for execution by the at least one processor, the one or more programs including instructions for:

obtaining words from a first user's set of one or more databases;

processing, using a set of one or more pronunciation guessers, the words from the first user's set of one or more databases, the processing producing additional phonetic data derived from the words, the additional phonetic data forming an extended phonetic dictionary for the first user;

receiving a speech input;

processing the speech input by comparing phonemes detected in the speech input to the phonetic dictionary and to the extended phonetic dictionary to determine a set of one or more matches; and

determining a best match from the set of one or more matches.

19. The electronic device of claim 18, wherein the first user's set of one or more databases comprises a contacts database with names of people and telephone numbers, and wherein the phonetic dictionary is for natural language speech recognition, and wherein the speech input includes a request that a telephone call be placed to another user's device.

20. The electronic device of claim 19, further comprising instructions for:

obtaining changes in the contacts database made as a result of the production of the additional phonetic data; and

processing, using the set of one or more pronunciation guessers, the changes to update the extended phonetic dictionary based on the changes, wherein the obtaining of the changes occurs in response to the changes being made.

21. The electronic device of any of claims 18-20, wherein the set of one or more pronunciation guessers comprise pronunciation guessers for a plurality of locales, each locale having its own pronunciation guesser.

22. The electronic device of any of claims 18-20, wherein the set of one or more pronunciation guessers comprises different pronunciation guessers for a same locale.

23. The electronic device of any of claims 18-22, further comprising instructions for:  
determining automatically a set of languages from the words in the first user's set of one or more databases; and

selecting a subset of the set of one or more pronunciation guessers, the subset selected based upon the set of languages that were automatically determined from the words in the first user's set of one or more databases;

wherein the processing which produces the additional phonetic data uses only the subset of the set of one or more pronunciation guessers when producing the additional phonetic data.

24. The electronic device of any of claims 19-23, wherein the electronic device is a server that is coupled through a wireless network to a device of the first user which includes the contacts database, and wherein the server obtains the words in the contacts database from the device of the first user through the wireless network, and wherein the server receives the speech input from the device of the first user through the wireless network, and wherein the server transmits the best match to the device of the first user through the wireless network.

25. The electronic device of claim 24, wherein the server executes the one or more programs for a plurality of users, and each user in the plurality of users has a dedicated extended phonetic dictionary that is unique to each user.

26. The electronic device of any of claims 19-23, wherein the electronic device is the device of the first user which includes the contacts database.

27. An electronic device, comprising:  
means for storing a phonetic dictionary for speech recognition;  
means for obtaining words from a first user's set of one or more databases;  
means for processing, using a set of one or more pronunciation guessers, the words from the first user's set of one or more databases, the processing producing additional phonetic data derived from the words, the additional phonetic data forming an extended phonetic dictionary for the first user;  
means for receiving a speech input;  
means for processing the speech input by comparing phonemes detected in the speech input to the phonetic dictionary and to the extended phonetic dictionary to determine a set of one or more matches; and  
means for determining a best match from the set of one or more matches.
28. An information processing apparatus for use in an electronic device, comprising:  
means for storing a phonetic dictionary for speech recognition;  
means for obtaining words from a first user's set of one or more databases;  
means for processing, using a set of one or more pronunciation guessers, the words from the first user's set of one or more databases, the processing producing additional phonetic data derived from the words, the additional phonetic data forming an extended phonetic dictionary for the first user;  
means for receiving a speech input;  
means for processing the speech input by comparing phonemes detected in the speech input to the phonetic dictionary and to the extended phonetic dictionary to determine a set of one or more matches; and  
means for determining a best match from the set of one or more matches.
29. A method comprising features of any combination of the methods of claims 10-17.
30. A system comprising one or more electronic devices, each comprising one or more of:  
an input device for receiving speech inputs;  
one or more processors;  
memory; and  
one or more programs, wherein the one or more programs are stored in the memory and configured to be executed by the one or more processors, the one or more programs including instructions for performing any of the methods of claims 10-17.

31. A computer readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with an input device for receiving speech inputs, cause the device to perform any of the methods of claims 10-17.

32. An electronic device, comprising:  
an input device for receiving speech inputs; and  
means for performing any of the methods of claims 10-17.

33. An information processing apparatus for use in an electronic device with an input device for receiving speech inputs, comprising:  
means for performing any of the methods of claims 10-17.

34. An electronic device, comprising:  
a speech receiving unit configured to receive a speech input;  
a user database unit configured to store a first user's set of one or more databases;  
a phonetic dictionary storage unit configured to store a phonetic dictionary for speech recognition; and  
a processing unit coupled to the speech receiving unit and the phonetic dictionary storage unit, the processing unit configured to:  
obtain words from the first user's set of one or more databases;  
process, using a set of one or more pronunciation guessers, the words from the first user's set of one or more databases, the processing producing additional phonetic data derived from the words, the additional phonetic data forming an extended phonetic dictionary for the first user;  
acquire a speech input from the speech receiving unit;  
process the speech input by comparing phonemes detected in the speech input to the phonetic dictionary and to the extended phonetic dictionary to determine a set of one or more matches; and  
determine a best match from the set of one or more matches.

35. The electronic device of claim 34, wherein the first user's set of one or more databases comprises a contacts database with names of people and telephone numbers, and wherein the phonetic dictionary is for natural language speech recognition, and wherein the speech input includes a request that a telephone call be placed to another user's device.

36. The electronic device of claim 35, wherein the processing unit is further configured to:

obtain changes in the contacts database made as a result of the production of the additional phonetic data; and

process, using the set of one or more pronunciation guessers, the changes to update the extended phonetic dictionary based on the changes, wherein the obtaining of the changes occurs in response to the changes being made.

37. The electronic device of any of claims 34-36, wherein the set of one or more pronunciation guessers comprise pronunciation guessers for a plurality of locales, each locale having its own pronunciation guesser.

38. The electronic device of any of claims 34-36, wherein the set of one or more pronunciation guessers comprises different pronunciation guessers for a same locale.

39. The electronic device of any of claims 34-38, wherein the processing unit is further configured to:

determine, automatically, a set of languages from the words in the first user's set of one or more databases; and

select a subset of the set of one or more pronunciation guessers, the subset selected based upon the set of languages that were automatically determined from the words in the first user's set of one or more databases;

wherein the processing which produces the additional phonetic data uses only the subset of the set of one or more pronunciation guessers when producing the additional phonetic data.

40. The electronic device of any of claims 35-39, wherein the electronic device is the device of the first user which includes the contacts database.

FIG. 1

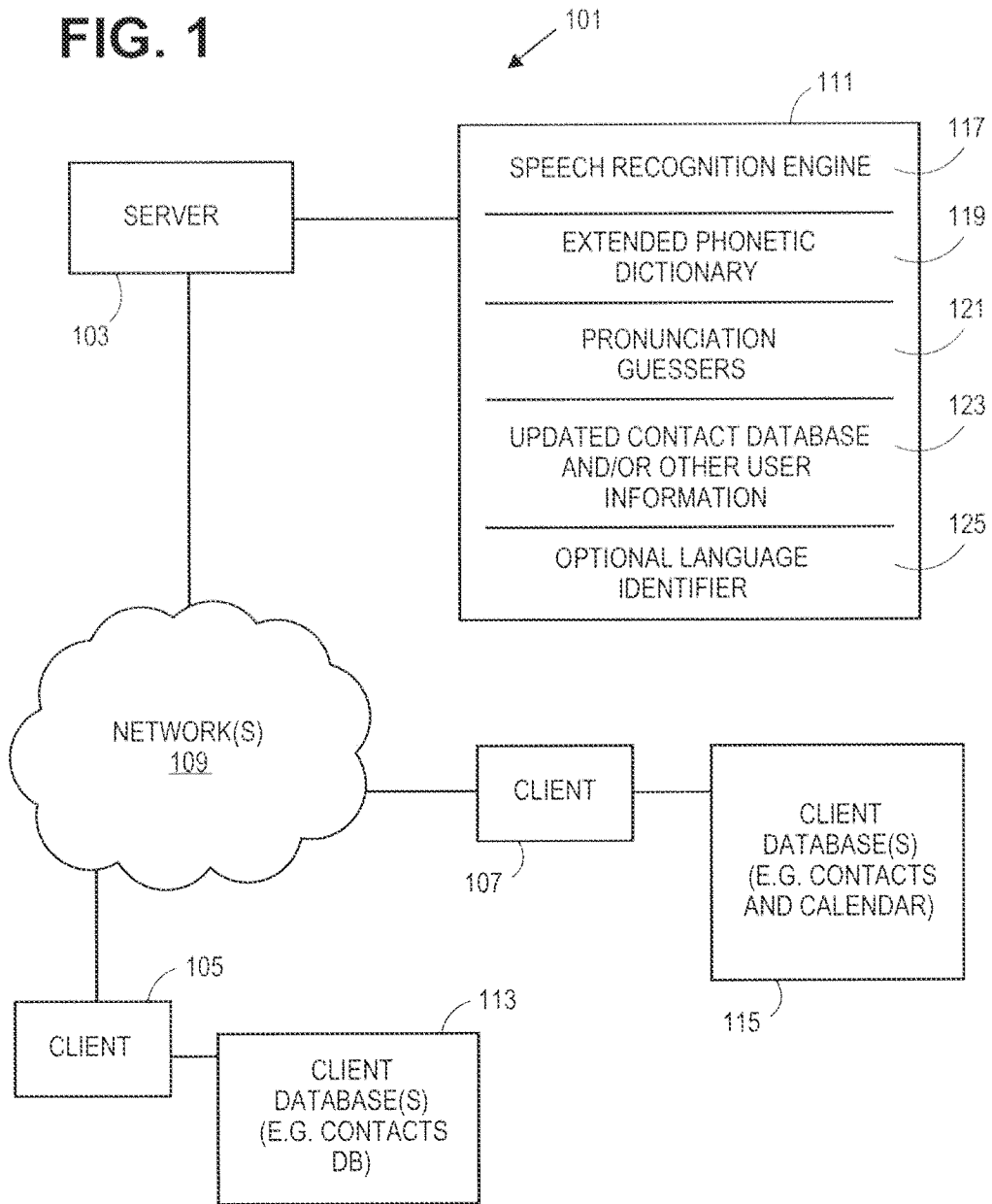
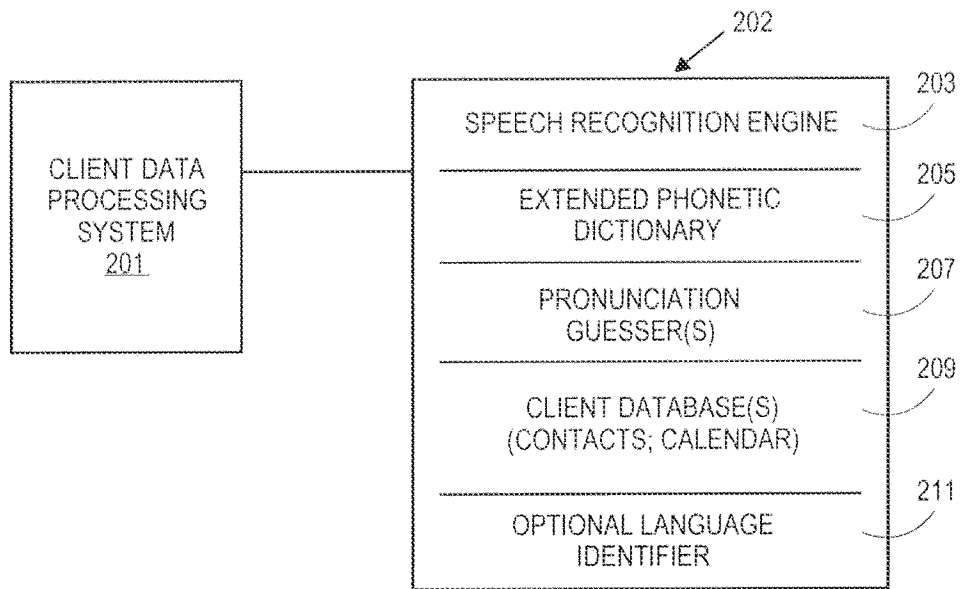
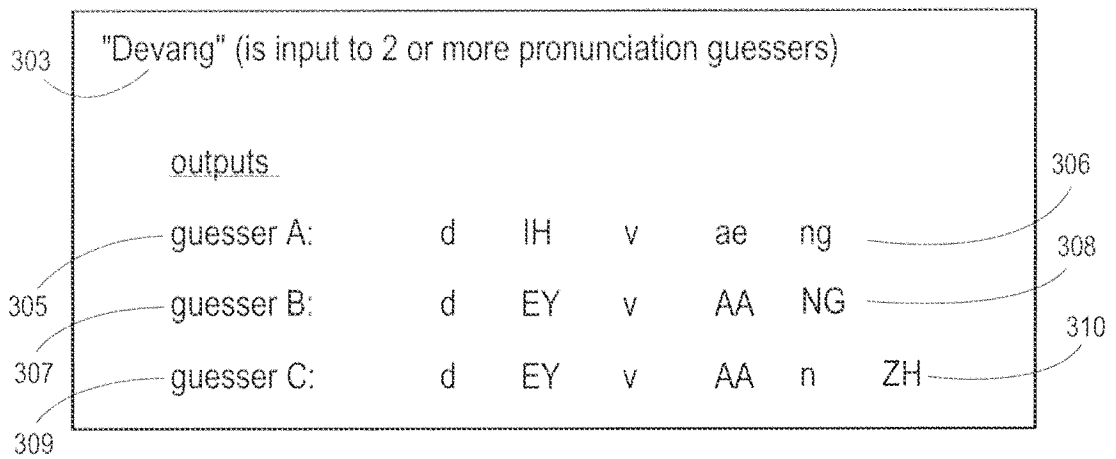


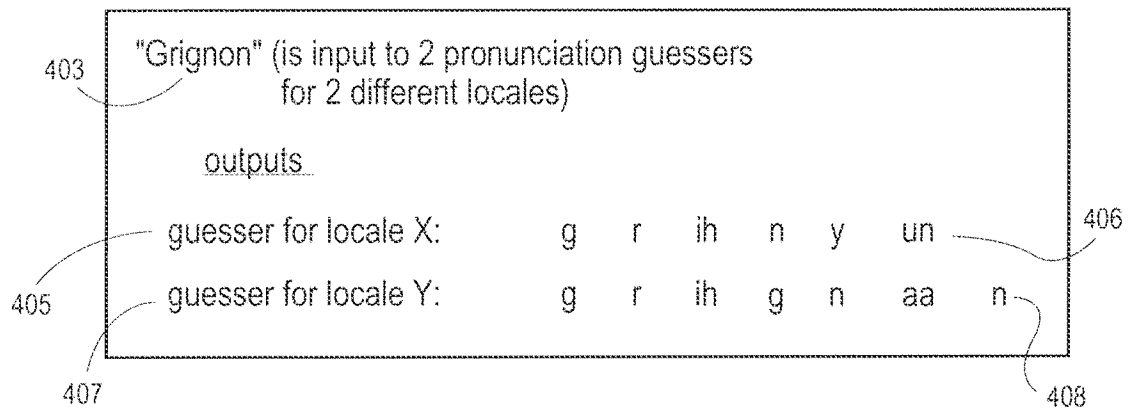
FIG. 2



**FIG. 3**



**FIG. 4**





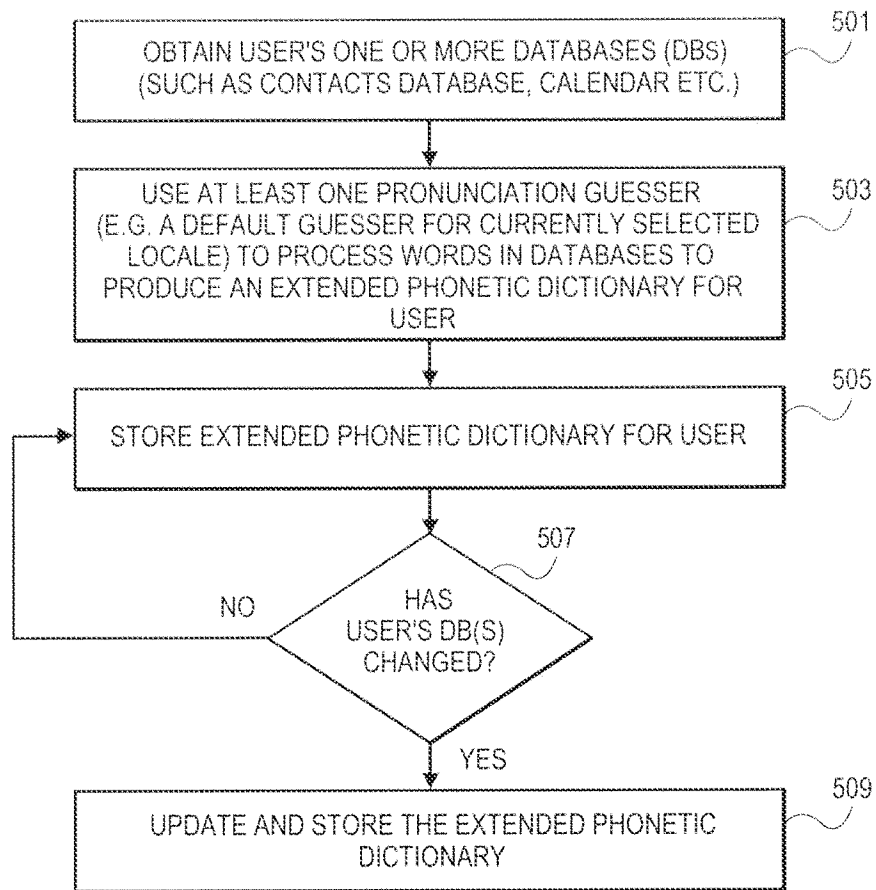


FIG. 5

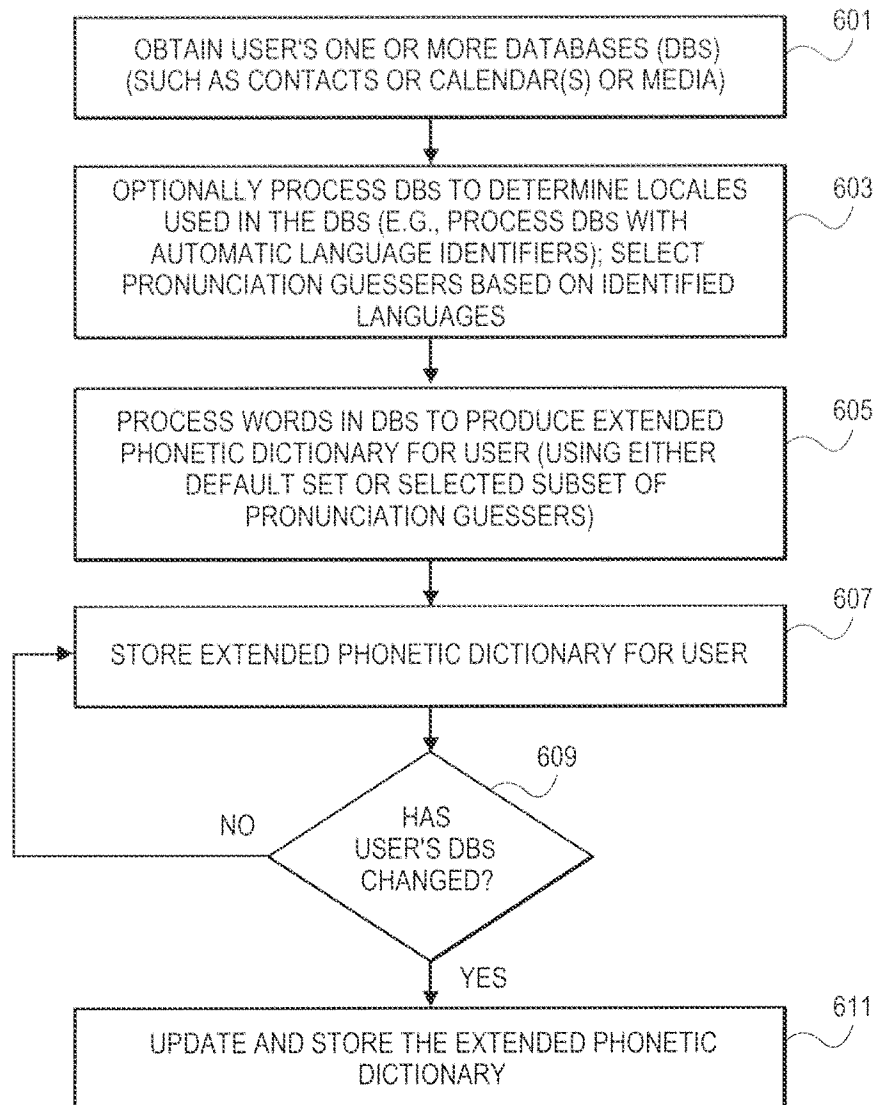
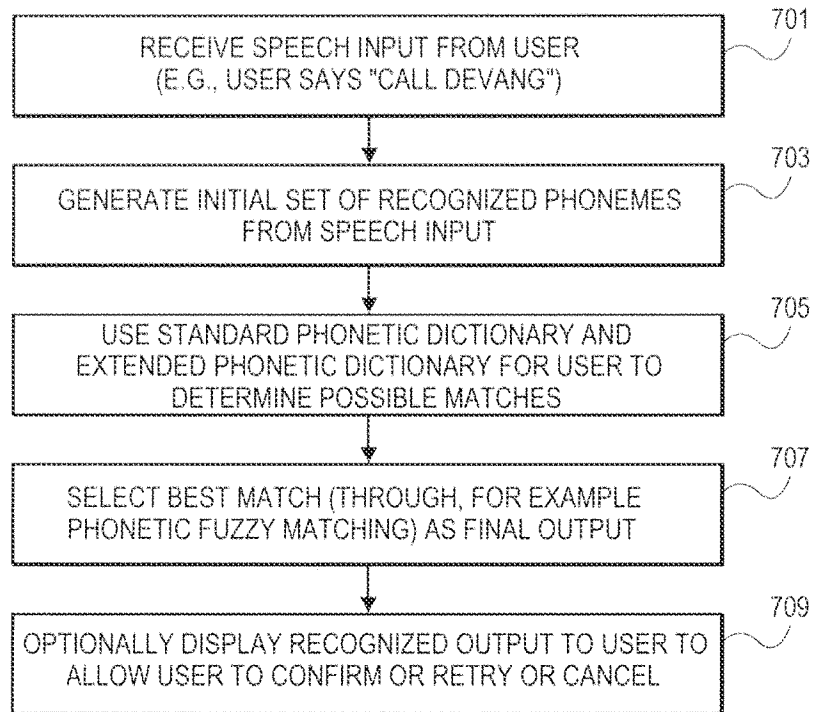
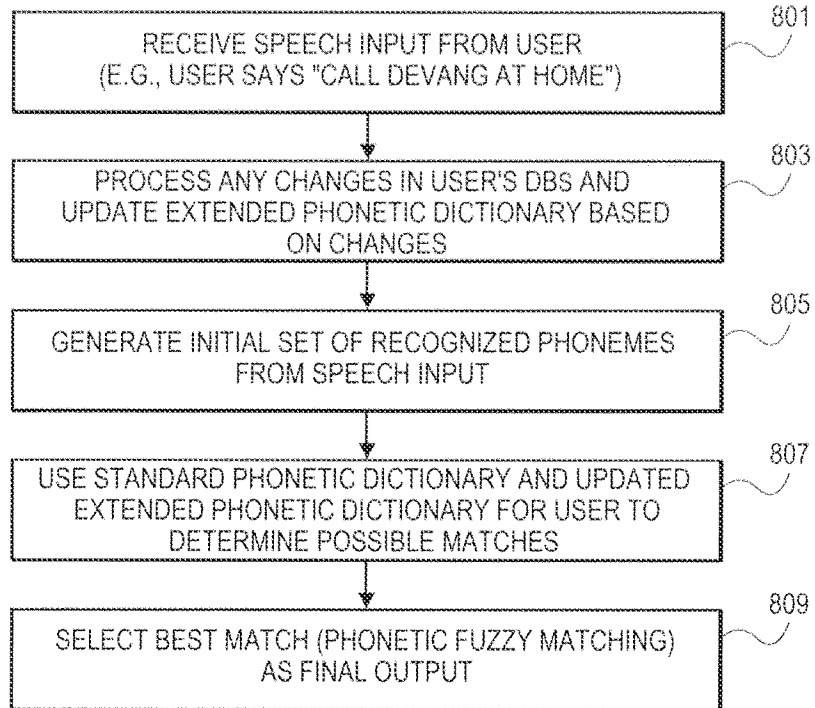


FIG. 6

**FIG. 7**



**FIG. 8**

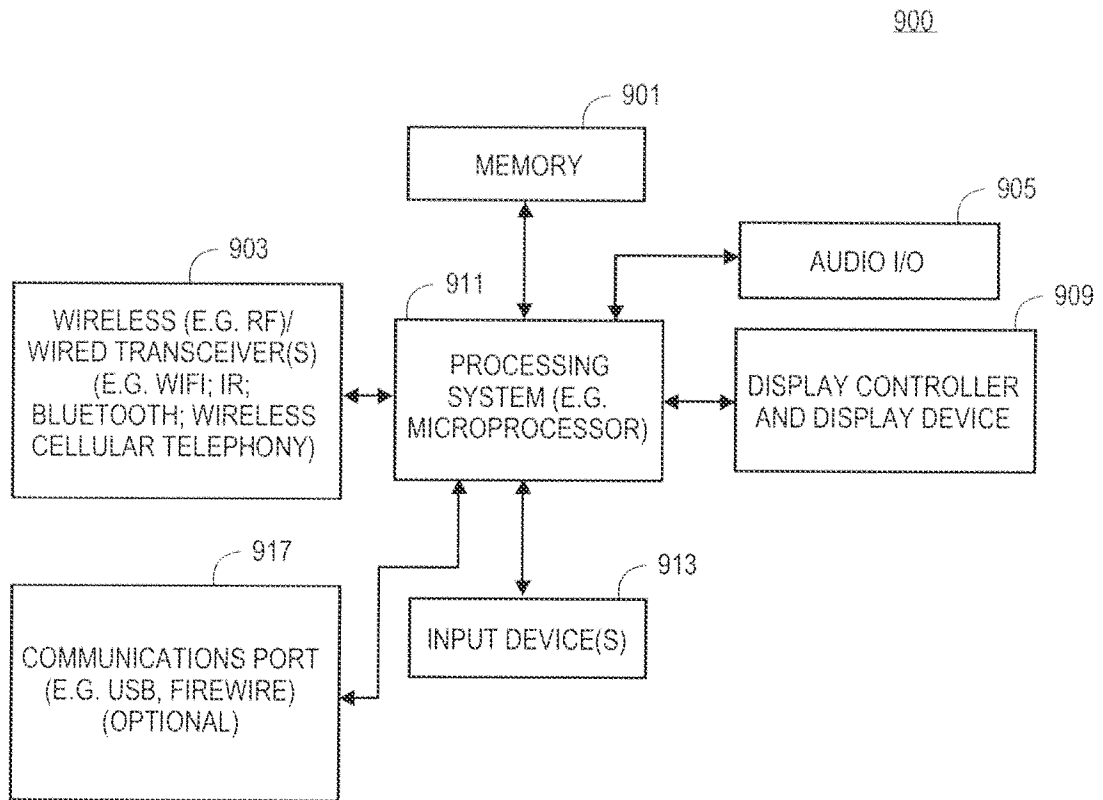


FIG. 9

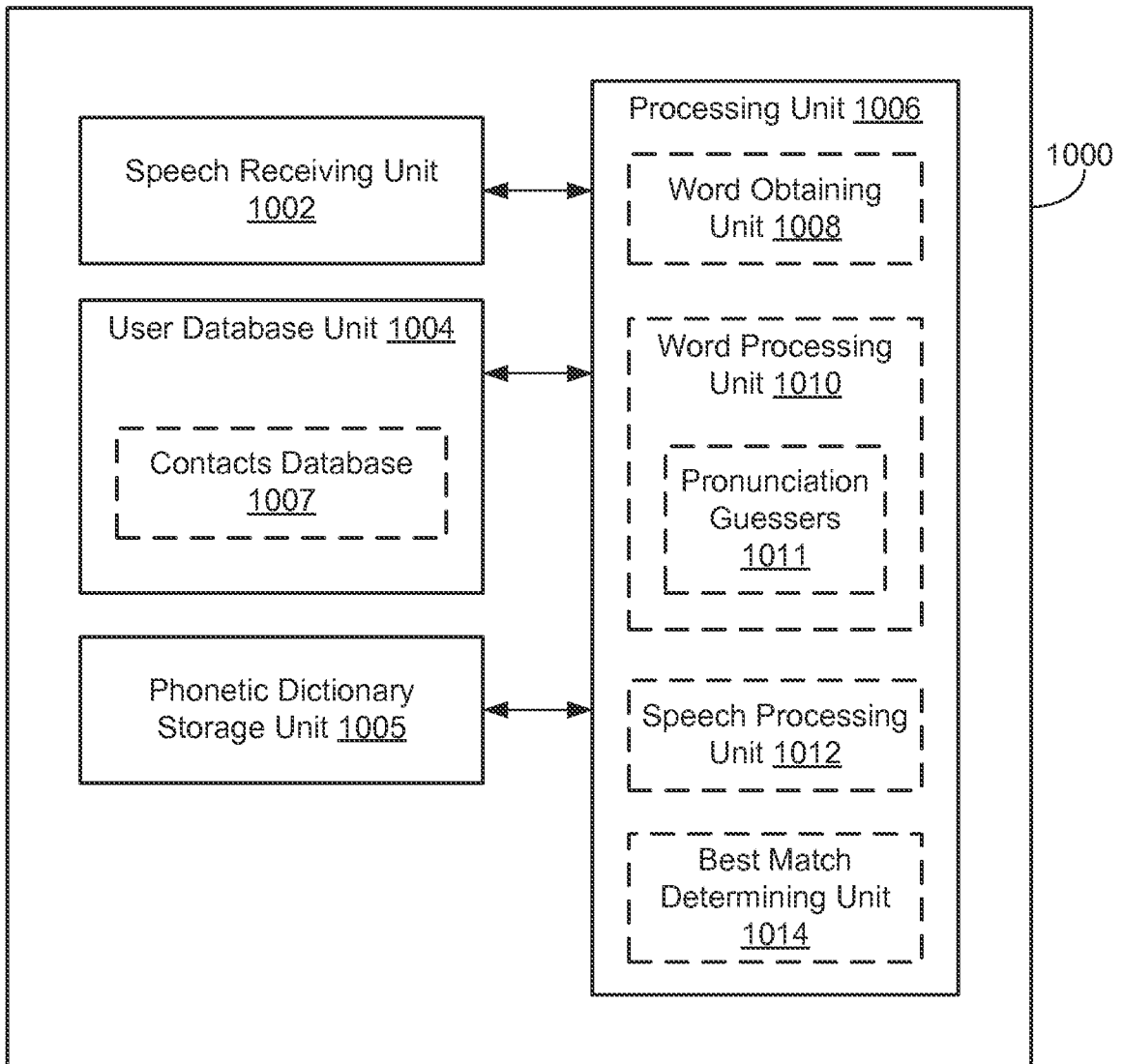


Figure 10

INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2013/044574

A. CLASSIFICATION OF SUBJECT MATTER  
INV. G10L15/187  
ADD. G10L15/06

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
G10L

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2006/173683 A1 (ROTH DANIEL L [US] ET AL) 3 August 2006 (2006-08-03)	1-3, 8-12, 17-20, 26-36,40
Y	paragraph [0003] paragraphs [0019] - [0023], [0024], [0028] paragraph [0030] paragraph [0042]  -----  -/--	4-7, 13-16, 21-25, 37-39

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search  18 September 2013	Date of mailing of the international search report  27/09/2013
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  Krembel, Luc

## INTERNATIONAL SEARCH REPORT

International application No

PCT/US2013/044574

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>VIIKKI I ET AL: "Speaker- and language-independent speech recognition in mobile communication systems", 2001 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. (ICASSP). SALT LAKE CITY, UT, MAY 7 - 11, 2001; [IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP)], NEW YORK, NY : IEEE, US, vol. 1, 7 May 2001 (2001-05-07), pages 5-8, XP010803072, DOI: 10.1109/ICASSP.2001.940753 ISBN: 978-0-7803-7041-8 paragraphs [0003], [03.2], [03.3] figure 1</p> <p style="text-align: center;">-----</p>	4,6,13, 15,21, 23,37,39
Y	<p>DENIS JOUVET ET AL: "Evaluating grapheme-to-phoneme converters in automatic speech recognition context", 2012 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP 2012) : KYOTO, JAPAN, 25 - 30 MARCH 2012 ; [PROCEEDINGS], IEEE, PISCATAWAY, NJ, 25 March 2012 (2012-03-25), pages 4821-4824, XP032228234, DOI: 10.1109/ICASSP.2012.6288998 ISBN: 978-1-4673-0045-2 paragraph [0006]</p> <p style="text-align: center;">-----</p>	5,14,22, 38
Y	<p>US 2011/054894 A1 (PHILLIPS MICHAEL S [US] ET AL) 3 March 2011 (2011-03-03) abstract figures 1,2 paragraphs [0047], [0070]</p> <p style="text-align: center;">-----</p>	7,16,24, 25
X	<p>EP 1 291 848 A2 (NOKIA CORP [FI]) 12 March 2003 (2003-03-12)</p> <p>figures 2,4 paragraphs [0002], [0006] paragraph [0009] paragraph [0028] paragraph [0037] paragraph [0043]</p> <p style="text-align: center;">-----</p>	1,10,18, 27,28, 30,34
X	<p>US 2004/049388 A1 (ROTH DANIEL L [US] ET AL) 11 March 2004 (2004-03-11)</p> <p>figure 10 paragraph [0123] paragraph [0126] paragraphs [0155], [0156] paragraph [0344]</p> <p style="text-align: center;">-----</p>	1,10,18, 27,28, 30,34



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2013/044574

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006173683	A1	03-08-2006	CN 101164102 A
			EP 1844464 A2
			JP 2008529101 A
			JP 2013047809 A
			KR 20070100837 A
			US 2006173683 A1
			WO 2006084144 A2
-----			
US 2011054894	A1	03-03-2011	NONE
-----			
EP 1291848	A2	12-03-2003	EP 1291848 A2
			US 2003050779 A1
-----			
US 2004049388	A1	11-03-2004	US 2004049388 A1
			WO 2004023455 A2
-----			