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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 December 2005 (01.12.2005)

PCT

(10) International Publication Number
WO 2005/114652 A1

(51) International Patent Classification[7]: G10L 13/08, 15/06

(21) International Application Number: PCT/EP2005/005568

(22) International Filing Date: 23 May 2005 (23.05.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
04012134.5 21 May 2004 (21.05.2004) EP

(71) Applicant *(for all designated States except US)*: **HARMAN BECKER AUTOMOTIVE SYSTEMS GMBH** [DE/DE]; Becker-Goering-Str. 16, 76307 Karlsbad (DE).

(72) Inventors; and

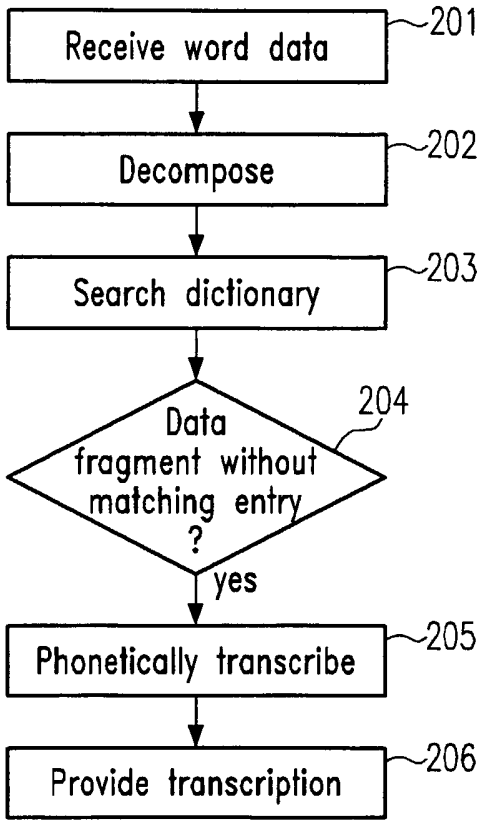
(75) Inventors/Applicants *(for US only)*: **SCHULZ, Matthias** [DE/DE]; Silcherweg 39, 89198 Westerstetten (DE). **GERL, Franz** [DE/DE]; Thalfinger Str. 62, 89233 Neu-Ulm (DE). **SCHWARZ, Markus** [DE/DE]; Virchowstr. 5, 89075 Ulm (DE). **KOSMALA, Andreas** [DE/DE]; Am Krummbach 10, 88471 Laupheim (DE). **JESCHKE, Bärbel** [DE/DE]; Finkenweg 3, 89173 Lonsee (DE).

(74) Agent: **WEIGELT, Udo**; Grünecker, Kinkeldey, Stockmair & Schwanhäusser, Maximilianstrasse 58, 80538 München (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[Continued on next page]

(54) Title: SPEECH RECOGNITION SYSTEM



(57) Abstract: The invention is directed to a method for automatically generating a vocabulary for a speech recognizer, comprising the steps of receiving digital word data, in particular, name data, automatically searching for a word entry in a predetermined dictionary matching at least part of the received word data, wherein the dictionary comprises a phonetic transcription for each word entry and automatically phonetically transcribing each part of the received word data for which no matching word entry was determined.

WO 2005/114652 A1

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

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

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

AUTOMATIC WORD PRONUNCIATION GENERATION FOR SPEECH RECOGNITION

The invention is directed to a speech recognition system and to a method for automatically generating a vocabulary for a speech recognizer.

Applications for speech recognition systems can be found in many different fields. For example, speech recognition systems can be used to generate texts. In this case, a text is spoken (dictated) by a user and transcribed into written text by the speech recognition system.

In many other fields, speech dialog systems are used to provide a comfortable and natural interaction between a human user and a machine. Depending on the device the user wants to interact with, a speech dialog system can be used to enable the user to get information, to order something or control the device in some other ways. For example, a speech dialog system can be employed in a car to allow the user control of different devices such as a mobile phone, car radio, navigation system and/or air condition system.

There are different types of speech recognizers that can be utilized. For example, the speech recognizer can be an isolated word recognizer or a compound word recognizer. In the case of the first one, a user has to separate subsequent input words by sufficiently long pauses such that the system can determine the beginning and the end of the word. In the case of the latter, on the other hand, the beginning and end of words are detected by the compound word recognizer itself that allows the user to speak in a more natural way.

A speech recognizer is based on the principle that a speech input is compared with previously stored speech patterns of a dictionary. If an input word is "sufficiently" similar (according to a given criterion or distance measure) to a stored speech pattern, this word is recognized accordingly. Speech recognition algorithms can be based on different methods such as template matching, Hidden Markov Models (HMM) and/or artificial neural networks.

In many different applications, speech recognition systems are used as embedded systems in the context of other devices being connected thereto. Typical examples for such embedded systems are mobile phones, PDA's, or car entertainment and control systems comprising a car radio/CD player and a navigation system.

Particularly in the case of embedded systems, the CPU power and memory provided for the speech recognition part is limited. On the other hand, in different applications, a large vocabulary and computing power is necessary to recognize different proper names. This is the case, for example, if a speech recognizer is used to access an entry of a telephone directory of a mobile phone and initiate calling only via speech commands, or when a car radio is to be controlled by selecting radio stations by speaking their names.

In view of this, it is the problem underlying the invention to provide a speech recognition system and a method that allows to operate a speech recognizer with reduced memory and computing power requirements.

This problem is solved by the method according to claim 1 and the speech recognition system according to claim 13.

Accordingly, a method for automatically generating a vocabulary for a speech recognizer is provided, comprising the steps of:

a) receiving digital word data, in particular, name data,

b) automatically searching for a word entry in a predetermined dictionary matching at least part of the received word data, wherein the dictionary comprises a phonetic transcription for each word entry,

c) automatically phonetically transcribing each part of the received word data for which no matching word entry was determined.

Here and in the following, the term "word" or "word data" denotes a sequence of characters wherein the term "character" comprises letters, numbers (digits) and special characters such as a dash, a blank, or a dot. Different alphanumeric characters can be separated by a blank, for example. The word data are received in digital form. Phonetically transcribing a word or a part of a word results in a speech pattern corresponding to how the word or part of a word is pronounced.

Compared to the case of prior art, this method enables to perform speech recognition without requiring to store a large dictionary comprising a huge number of entries covering all possible speech inputs in advance. The method allows to dynamically generate a vocabulary for a speech recognizer upon receipt of corresponding word data. In this way, the speech patterns for the speech recognizer are much better adapted to current conditions and requirements of the speech recognizer and its user during operation. In particular, if word data is received, it is checked whether the predetermined dictionary already comprises entries matching the whole received word data or parts thereof. If this is the case, a corresponding phonetic transcription of the word or parts thereof is already present. If not, however, the remaining parts of the received word data (which can also be all of the received word data) are phonetically transcribed such that corresponding speech patterns are obtained.

Step b) can comprise decomposing the word data into data fragments according to predetermined categories and searching for a fragment entry in the dictionary matching a data fragment or matching a sequence of data fragments, and step c) can comprise phonetically transcribing each data fragment for which no matching fragment entry was determined.

A data fragment consists of one character or is a sequence (string) of several characters (such as letters, numbers or special characters). A fragment entry of the dictionary can match a single data fragment or a sequence of data fragments of the word data, wherein the data fragments of such a sequence need not belong to the same category. It is possible to only search a part of the dic-

tionary corresponding to the category of the data fragment for which a matching fragment entry is to be determined. Thus, a dictionary can be provided in which each entry is associated to a predetermined category. In other words, the dictionary's entries can be categorized such that the dictionary comprises sub-dictionaries the entries of which belong to a common predetermined category, respectively. This reduces the required computing power and computing time since, for each data fragment, only parts of the dictionary corresponding to the category of the data fragments are to be searched.

The decomposing step can comprise decomposing the word data into data fragments solely consisting of letters, data fragments solely consisting of numbers, and/or data fragments consisting solely of special characters.

With these categories, letter fragments, number fragments, and special character fragments are obtained, each fragment consisting of one character or a sequence of characters all belonging to only one of these categories, respectively. Such a categorization is particularly useful in the context of address books or radio stations, since, in these cases, the corresponding word data usually comprises both letters and numbers.

A data fragment can consist of a string of characters appearing successively in the word data string. The characters of a data fragment, however, can be separated by a blank, a dash or another special character. Preferably, word data comprising two sequences of alphanumeric characters being separated by at least one special character can be decomposed into two fragments each corresponding to a sequence of alphanumeric characters and into a fragment consisting of the at least one special character. Alphanumeric characters are either letters or numbers. For example, if two sequences of letters of the word data are separated by a blank, two fragments are formed, each corresponding to one of the two letter fragments; a third fragment consists of the blank. Since in many cases, sequences of alphanumeric characters of the word data being separated by a dash or a blank may constitute separate entities (such as first name and family name), this yields an advantageous decomposition of the

word data for further processing. It is possible to perform additional decompositions according to other categories (letter category, number category, for example); and the order of the decompositions can also be exchanged.

Step c) can comprise determining according to a predetermined criterion whether to phonetically transcribe a part of the received word data in spelled form, in pronounced form, or in composed spelled and pronounced form.

Spelled form yields a phonetic transcription in which each character of the fragment is spelled, whereas in pronounced from, a sequence of characters is pronounced or enunciated as a whole word composed of the characters. In composed form, a part of the data fragment is spelled and another part is pronounced.

The predetermined criterion may depend on the length of the data fragment (number of characters), neighboring fragments, upper and lower case characters and/or the appearance of consonants and vowels.

In particular, each letter of a part of the received word data solely consisting of consonants can be transcribed in spelled form. This yields a phonetic transcription of the letter sequence of a data fragment that most likely would correspond to the pronunciation, i.e. spelling, of the data fragment by a speaker.

Step a) can comprise receiving digital word data via Radio Data System (RDS) or Radio Broadcast Data System (RBDS).

RDS or RBDS allows to also transmit digital data (such as the name of the radio station or other information) in addition to the audio signal itself. Hence, this method can be advantageously used for a speech recognizer for a radio, e.g. a car radio, in particular, for controlling a radio via speech commands. Since the radio signals of RDS or RBDS also comprise digital word data, these word data are preferably used to generate a corresponding vocabulary for the speech recognizer.

Step a) can be preceded by automatically requesting digital word data. For example, the requesting step can comprise initiating scanning a radio frequency band and searching for receivable radio stations. In this way, the dynamically generated vocabulary for a speech recognizer to control a radio is always adapted to current circumstances, namely the receivable radio stations; a vocabulary obtained in this way comprises the necessary data that can currently be used.

Step a) can comprise uploading word data from a name database, in particular, an address book database or a telephone directory database. In this way, the name database comprising a set of proper names allows to specifically enlarge the vocabulary for a speech recognizer such that these proper names have an increased recognition probability by the speech recognizer. Other name databases are possible as well; for example, a name database can comprise a list of the names of radio stations.

In the above-described methods, a dictionary can be used that further comprises a synonym and/or an abbreviation and/or at least two phonetic transcriptions for at least one of the word entries.

In this way, a user has the possibility to input a speech command in different ways. For example, the user can enunciate a term as a whole or use its abbreviation; if both are in the dictionary, both alternatives would be recognized. In some cases, it could also be possible that a term can be pronounced in different ways, for example, using an English or a German pronunciation; also these different pronunciations would be recognized if both phonetic transcriptions are part of the dictionary.

The above-described method can further comprise the step of providing the speech recognizer with each phonetically transcribed part and the corresponding part of the received word data. Thus, the vocabulary of the speech recognizer is enlarged and the functionality of the speech recognizer improved.

Steps b) and c) can be repeated according to a predetermined criterion, in particular, after predetermined time intervals and/or each time word data is received during operating time of the speech recognizer. Also the step of providing the speech recognizer with each phonetically transcribed part can be repeated according to the predetermined criterion.

This allows to regularly update the vocabulary. For example, in case of receiving digital word data via RDS or RBDS, the steps can be repeated each time a new radio station is receivable and/or each time a radio station is no longer receivable. Thus, the above described methods are particularly useful for a speech recognizer in a vehicular cabin.

The term "operating time" designates the time in which the speech recognizer can be used (even if a push-to-talk (PTT) key is still to be pressed) and not only the time during which the speech recognizer is active, i.e. is actually processing speech input after a PTT key was pressed.

The invention also provides a computer program product comprising one or more computer readable media having computer-executable instructions for performing the steps of one of the previously described methods.

The invention also provides a speech recognition system, comprising:

a speech recognizer for recognizing speech input,

an interface configured to receive word data, in particular, name data,

a memory in which a predetermined dictionary is stored, the dictionary comprising digital word entries and a phonetic transcription for each word entry,

search means configured to automatically search for a word entry in a dictionary matching at least part of the received word data,

transcription means configured to automatically phonetically transcribe each part of received word data for which no matching word entry was determined.

With such a speech recognition system, the above-described methods can be performed in an advantageous way. In particular, the vocabulary for the speech recognizer can be dynamically generated and updated. It is to be understood that the different parts of the speech recognition system can be arranged in different possible ways. For example, the memory in which the predetermined dictionary is stored can be provided as a separate entity, or the speech recognizer can comprise this memory.

The search means can be configured to decompose the word data into data fragments according to predetermined categories and to search for a fragment entry in the dictionary matching a data fragment or matching a sequence of data fragments, and the transcription means can be configured to phonetically transcribe each data fragment for which no matching fragment entry was determined.

The search means can be configured to decompose the word data into data fragments solely consisting of letters, data fragments solely consisting of numbers, and/or data fragments solely consisting of special characters.

The search means can be configured such that word data comprising two sequences of alphanumeric characters being separated by at least one special character can be decomposed into two fragments each corresponding to a sequence of alphanumeric characters and into a fragment consisting of the at least one special character.

The transcription means can be configured to determine according to a predetermined criterion whether to phonetically transcribe a part of the received word data in spelled form, in pronounced form, or in composed spelled and pronounced form.

The transcription means can be configured to transcribe each letter of a part of the received word data solely consisting of consonants in spelled form.

The interface can be configured to receive digital word data via Radio Data System (RDS) or Radio Broadcast Data System (RBDS).

The interface can be configured to automatically request digital word data.

The interface can be configured to upload word data from a name database, in particular, an address book database or a telephone directory database.

The dictionary can further comprise a synonym and/or an abbreviation and/or at least two phonetic transcriptions for at least one of the word entries.

The transcription means can be configured to store each phonetically transcribed part and the corresponding part of the received word data in a memory being accessible by the speech recognizer.

The search means and the transcription means can be configured to repeat searching and transcribing according to a predetermined criterion, in particular, after predetermined time intervals and/or each time word data is received during operating time of the speech recognizer.

Further features and advantages will be described in the following with respect to the drawings.

Fig. 1 illustrates an example of a speech recognition system for controlling a radio;

Fig. 2 illustrates the course of action of an example of a method for generating a vocabulary for a speech recognizer;

Fig. 3 illustrates the course of action of an example of generating a radio station vocabulary for a speech recognizer;

Fig. 4 illustrates an example of a list of radio stations; and

Fig. 5 illustrates an example of dictionary entries for radio stations.

While the present invention is described in the following with respect to different embodiments and figures, it should be understood that the following detailed description as well as the drawings are not intended to limit the present invention to the particular illustrative embodiments disclosed, but rather that the described illustrative embodiments merely exemplify the various aspects of the present invention, the scope of which is defined by the appended claims.

In Fig. 1, the arrangement of different components of an example of a speech recognition system is shown. In particular, the speech recognition system comprises a speech recognizer 101 that is responsible for recognizing a speech input. Such a speech input is received by the speech recognizer 101 via a microphone 102.

In general, the speech recognizer can be connected to one microphone or to several microphones, particularly to a microphone array (not shown). If a microphone array were used, the signals emanating from such a microphone array would be processed by a beamformer to obtain a combined signal that has a specific directivity. This is particularly useful if the speech recognition system is to be used in a noisy environment such as in a vehicular cabin.

The speech recognition system as illustrated in Fig. 1 is intended to control a radio 103 via speech commands. The speech recognizer can be an isolated

word recognizer or a compound word recognizer. Its recognition algorithm may be based on template matching, Hidden Markov Models and/or artificial neuron networks, for example. In any of these cases, the speech recognizer 101 compares a speech input from a user with speech patterns that have been previously stored in a memory 104. If the speech input is sufficiently closed (according to a predetermined distance measure) to one of the stored speech patterns, the speech input is recognized as the speech pattern. In memory 104, a standard vocabulary is stored for controlling the functionality of connected devices such as the radio 103. However, other devices such as a navigation system or an air conditioning can also be connected to the speech recognizer.

For example, a user could speak the words "volume" and "up". If speech patterns corresponding to these two words are present in the speech recognizer, the command "volume up" would be recognized and a corresponding electronic signal be sent from the speech recognizer to the radio 103 so as to increase the volume.

In the case of the radio, particularly a car radio, one would also like to choose a radio station via corresponding speech command. However, the names of radio stations are often specific proper names that do not form part of a standard vocabulary of a speech recognizer. This is particularly due to limited memory.

Because of that, an interface 105 is provided which serves to receive the names of different radio stations in digital form. The interface is configured to receive the data via a suitable network, for example, from a radio station in wireless form.

In particular, the interface 105 can be configured to receive digital data via Radio Data System (RDS) or Radio Broadcast Data System (RBDS). According to these systems, signals received from a radio station not only contain audio data, but also additional information. For example, the station's name

can be simultaneously transmitted in digital form. Many RDS or RBDS receivers are capable of presenting the station's name on a display.

In the present case, this name information is used to provide additional vocabulary data for the speech recognizer. To achieve this, a search means 106 is connected to the interface 105. As will be described in more detail below, the search means 106 is responsible to determine whether at least parts of a name received by the interface 105 are present in a predetermined dictionary that has been stored in memory 107.

The memory 107 is not only connected to the search means 106, but also to the speech recognizer 101. In particular, the speech recognizer has also access to the dictionary stored in memory 107. In the example shown in Fig. 1, the memory 107 is shown as an additional part outside the speech recognizer. Other configurations, however, are also possible; for example, the memory could be part of the search means or of the speech recognizer.

Furthermore, a transcription means 108 is connected to the search means 106 and to the speech recognizer 101. If some or all parts of incoming name data were not found in the dictionary stored in memory 107, these parts are passed to the transcription means 108 where they are phonetically transcribed. The transcription means may also be configured to determined which kind of phonetic transcription is performed (spelled form, pronounced form, or composed spelled and pronounced form).

Each phonetic transcription is passed to the speech recognizer 101, where it is stored in memory 104 so that it can be used by the speech recognizer. In this way, additional vocabulary is dynamically generated for the speech recognizer.

Also in this case, the additional vocabulary can be stored in other ways. For example, the transcription means 108 can also comprise a memory where this additional vocabulary is to be stored. Alternatively, the phonetically tran-

scribed parts can be stored in memory 107 in addition to the dictionary already present. In this case, the transcription means 108 need not be directly connected to the speech recognizer 101; it has to be connected to the memory 107, either directly or via the search means 106, for example.

In the following, an example of how to generate a vocabulary will be described with reference to Fig. 2. The steps of the method will be explained along the lines of the examples of a speech recognition system for controlling a radio as illustrated in Fig. 1. It is to be understood, however, that the steps of the method apply to other examples such as mobile phones as well.

In a first step 201, digital word data is received. In the case of a speech recognition system for a radio, the word data are names of radio stations. However, if the method is used for a mobile phone, for example, the word data could comprise entries of an address book or a telephone directory.

As an example, the received word data may comprise "SWR 4 HN". This stands for the radio station "Südwestrundfunk 4 Heilbronn". When receiving the name of this radio station, the corresponding frequency on which these radio signals are transmitted is also known: It is the frequency of the signal with which the name information was received.

In the following step 202, the name data "SWR 4 HN" is decomposed according to predetermined categories, in the present example, according to the categories "letter" and "number". The string of characters can be analyzed as follows.

One starts with the first character which is "S" and belongs to the category "letter". The subsequent characters "W" and "R" belong to the same category. After these three letters, there is a blank followed by the character "4", which is a number. Therefore, the sequence of characters belonging to the same category, namely, the category "letters" is terminated and a first data fragment "SWR" is determined. The following blank constitutes a next fragment.

The number "4" is followed by a blank and, then, by the character "H", which is a letter. Hence, the sequence of characters belonging to the same category consists only of a single character, namely the number "4"; this is a further fragment, again followed by a blank resulting in the next fragment. After this, a last fragment consisting of the letters "H" and "N" is determined. As a result, the word data "SWR 4 HN" was decomposed into fragments "SWR", "4", and "HN" and into two special character fragments consisting of a blank, respectively.

Other variants of decomposing the name data is possible as well. For example, first, of all, the data can be decomposed in different parts that are separated from another by a blank or a special character such as a dash or a dot. After that, one can perform the decomposition into letters and numbers as described above. In the case of the above example "SWR 4 HN", decomposition into sequences of alphanumeric characters being separated by a blank would already yield the three fragments "SWR", "4" and "HN" and the two special character fragments. A further decomposition into letter fragments and number fragments would not modify this decomposition.

It is to be understood that the different steps of the decomposition can also be exchanged.

In the next step 203, a dictionary is searched to determine whether there are any entries matching one or a sequence of the data fragments. The dictionary can comprise the names or abbreviations of major radio stations. For each data fragment or possibly for a sequence of data fragments, the dictionary is searched. The dictionary can also be decomposed into different sub-dictionaries each comprising entries belonging to a specific category. For example, a sub-dictionary can comprise only those entries consisting solely of letters and another sub-dictionary can only comprise entries consisting solely of numbers. Then, only the letter sub-dictionary would be searched with respect to letter data fragments and only the number sub-dictionary would be

searched with regard to number data fragments. In this way, the processing time can be considerably reduced.

In the following step 204, it is checked whether there is any data fragment for which no matching fragment entry was found in the dictionary. If this is not the case, the method can be terminated since the complete name data is already present in the dictionary. As the dictionary comprises the phonetic transcription for each of its entries, the speech recognizer has all the necessary information for recognizing these fragments.

However, if there are one or several data fragments for which no matching entry has been found in the dictionary, the method proceeds with step 205. In this step, each of these data fragments is phonetically transcribed. This means, a speech pattern is generated corresponding to the pronunciation of the data fragment. For this, text to speech (TTS) synthesizers known in the art can be used. In this step, it is also decided according to a predetermined criterion what phonetic transcription is to be performed. For example, according to such a criterion, for data fragments consisting of less than a predetermined number of characters, a phonetic transcription in spelled form is always to be chosen. The criterion can also depend (additionally or alternatively) on the appearance upper and lower case characters, on neighboring (preceding or following) fragments, etc.

If a letter data fragment only consists of consonants, it is advantageous to phonetically transcribe this fragment in spelled form as well. In other words, the resulting phonetic pattern corresponds to spelling the letters of the data fragment. This is particularly useful for abbreviations not containing any vowels which would also be spelled by a user. However, in other cases, it might be useful to perform a composed phonetic transcription consisting of phonetic transcriptions in spelled and in pronounced form.

In the last step 206, the phonetic transcriptions and the corresponding word data fragments are provided to the speech recognizer, for example, by passing

the phonetic transcription together with the corresponding data fragments to the memory of the speech recognizer and storing them in the memory. Thus, the vocabulary for speech recognition is extended.

Fig. 3 illustrates an example of how the word data to be received can be obtained in the case of a radio, particularly a car radio. According to the first step 301, the radio frequency band is scanned. This can be performed upon a corresponding request by a speech recognizer. During this scanning of the frequency band, all frequencies are determined at which radio signals are received.

In the following step 302, a list of receivable stations is determined. When scanning a frequency band, each time a frequency is encountered at which a radio signal is received, this frequency is stored together with the name of the corresponding radio station that is received as a digital signal together with the audio signal.

An example of a resulting list of receivable radio stations is shown in Fig. 4. The left column is the name of the radio station as received via RDS or RBDS and the right column lists the corresponding frequencies at which these radio stations can be received. It is to be understood that such a list can be stored in different ways.

In the next step 303, it is checked whether there is already a list of receivable radio stations present or whether the current list is amended with respect to a previously stored list of radio stations. The latter can happen in the case of a car radio, for example, when the driver is moving between different transmitter coverage areas, some radio stations may become receivable at a certain time whereas other radio stations may no longer be receivable.

If there is no previously stored radio station list present or if the list of receivable radio stations has been changed, the method proceeds to step 304 in which vocabulary corresponding to the list of radio stations is generated. This

would be done according to the method illustrated in Fig. 2. Otherwise, the method returns to step 301.

These steps can be performed continuously or regularly after predetermined time intervals.

Fig. 5 illustrates a dictionary that is searched in step 203 of the method shown in Fig. 2. To each fragment entry of the dictionary, a corresponding phonetic transcription is associated. For example, one entry can read "SWR". This entry is an abbreviation. For this entry, the dictionary also comprises the corresponding full word "Südwestrundfunk" together with its phonetic transcription. If there is a radio station called "Radio Energy", the dictionary could also comprise the fragment "Energy". For this entry, two different phonetic transcriptions are present, the first phonetic transcription corresponding to the (correct) English pronunciation and a second phonetic transcription corresponding to a German pronunciation of the word "energy". Thus, the term "energy" can be recognized even if a speaker uses a German pronunciation.

In the case of radio stations that are identified by their frequency, the dictionary can also comprise entries corresponding to different ways to pronounce or spell this frequency. For example, if a radio station is received at 94.3 MHz, the dictionary can comprise entries corresponding to "ninety-four dot three", "ninety-four three", "nine four three" etc. Therefore, a user may pronounce the "dot" or not; in both cases, the frequency is recognized.

In the foregoing, the method for generating a vocabulary for a speech recognizer was described in the context of a radio, in particular, a car radio. It is to be understood that this method can be used in other fields as well.

An example is the use of a speech recognizer for mobile phones. If a user wishes to control a mobile phone via speech commands, the speech recognizer has to recognize different proper names a user wishes to call. In such a case, a vocabulary can be generated based on an address book stored on the

SIM card of the mobile phone, for example. In such a case, this address book database is uploaded, for example, when switching on the mobile phone, and the method according to Fig. 2 can be performed. In other words, the steps of this method are performed for the different entries of the address book. Also in this case, a dictionary can be provided already comprising some names and their pronunciations. Furthermore, the dictionary can also comprise synonyms, abbreviations and/or different pronunciations for some or all of the entries. For example, the entry "Dad" can be associated to "Father" and "Daddy".

Further modifications and variations of the present invention will be apparent to those skilled in the art in view of this description. Accordingly, the description is to be construed as illustrative only and is for the purpose of teaching those skilled in the art the general manner of carrying out the present invention. It is to be understood that the forms of the invention as shown and described herein are to be taken as the presently preferred embodiments.

Claims

1. Method for automatically generating a vocabulary for a speech recognizer, comprising the steps of:

   a) receiving digital word data, in particular, name data,

   b) automatically searching for a word entry in a predetermined dictionary matching at least part of the received word data, wherein the dictionary comprises a phonetic transcription for each word entry,

   c) automatically phonetically transcribing each part of the received word data for which no matching word entry was determined.

2. Method according to claim 1, wherein step b) comprises decomposing the word data into data fragments according to predetermined categories and searching for a fragment entry in the dictionary matching a data fragment or matching a sequence of data fragments, and wherein step c) comprises phonetically transcribing each data fragment for which no matching fragment entry was determined.

3. Method according to claim 2, wherein the decomposing step comprises decomposing the word data into data fragments solely consisting of letters, data fragments solely consisting of numbers, and/or data fragments solely consisting of special characters.

4. Method according to claim 2 or 3, wherein word data comprising two sequences of alphanumeric characters being separated by at least one special character are decomposed into two fragments each corresponding to asequence of alphanumeric characters and into a fragment consisting of the at least one special character.

5. Method according to one of the preceding claims, wherein step c) comprises determining according to a predetermined criterion whether to phonetically transcribe a part of the received word data in spelled form, in pronounced form, or in composed spelled and pronounced form.

6. Method according to one of the preceding claims, wherein in step c), each letter of a part of the received word data solely consisting of consonants is transcribed in spelled form.

7. Method according to one of the preceding claims, wherein step a) comprises receiving digital word data via Radio Data System (RDS) or Radio Broadcast Data System (RBDS).

8. Method according to one of the preceding claims, wherein step a) is preceded by automatically requesting digital word data.

9. Method according to one of the preceding claims, wherein step a) comprises uploading word data from a name database, in particular, an address book database or a telephone directory database.

10. Method according to one of the preceding claims, wherein a dictionary is used that further comprises a synonym and/or an abbreviation and/or at least two phonetic transcriptions for at least one of the word entries.

11. Method according to one of the preceding claims, further comprising the step of providing the speech recognizer with each phonetically transcribed part and the corresponding part of the received word data.

12. Method according to one of the preceding claims, wherein steps b) and c) are repeated according to a predetermined criterion, in particular, af-

ter predetermined time intervals and/or each time word data is received during operating time of the speech recognizer.

13. Computer program product, comprising one or more computer readable media having computer-executable instructions for performing the steps of the method of one of the preceding claims.

14. Speech recognition system, comprising:

a speech recognizer for recognizing speech input,

an interface configured to receive word data, in particular, name data,

a memory in which a predetermined dictionary is stored, the dictionary comprising digital word entries and a phonetic transcription for each word entry,

search means configured to automatically search for a word entry in the dictionary matching at least part of the received word data,

transcription means configured to automatically phonetically transcribe each part of the received word data for which no matching word entry was determined.

15. Speech recognition system according to claim 14, wherein the search means is configured to decompose the word data into data fragments according to predetermined categories and to search for a fragment entry in the dictionary matching a data fragment or matching a sequence of data fragments, and wherein the transcription means is configured to phonetically transcribe each data fragment for which no matching fragment entry was determined.

16. Speech recognition system according to claim 15, wherein the search means is configured to decompose the word data into data fragments solely consisting of letters, data fragments solely consisting of numbers, and/or data fragments solely consisting of special characters.

17. Speech recognition system according to claim 15 or 16, wherein the search means is configured such that word data comprising two sequences of alphanumeric characters being separated by at least one special character are decomposed into two fragments each corresponding to a sequence of alphanumeric characters and into a fragment consisting of the at least one special character.

18. Speech recognition system according to one of the claims 14 – 17, wherein the transcription means is configured to determine according to a predetermined criterion whether to phonetically transcribe a part of the received word data in spelled form, in pronounced form, or in composed spelled and pronounced form.

19. Speech recognition system according to one of the claims 14 – 18, wherein the transcription means is configured to transcribe each letter of a part of the received word data solely consisting of consonants in spelled form.

20. Speech recognition system according to one of the claims 14 – 19, wherein the interface is configured to receive digital word data via Radio Data System (RDS) or Radio Broadcast Data System (RBDS).

21. Speech recognition system according to one of the claims 14 – 20, wherein the interface is configured to automatically request digital word data.

22. Speech recognition system according to one of the claims 14 – 21, wherein the interface is configure to upload word data from a name da-

tabase, in particular, an address book database or a telephone directory database.

23. Speech recognition system according to one of the claims 14 – 22, wherein the dictionary further comprises a synonym and/or an abbreviation and/or at least two phonetic transcriptions for at least one of the word entries.

24. Speech recognition system according to one of the claims 14 – 23, wherein the transcription means is configured to store each phonetically transcribed part and the corresponding part of the received word data in a memory being accessible by the speech recognizer.

25. Speech recognition system according to one of the claims 14 – 24, wherein the search means and the transcription means are configured to repeat searching and transcribing according to a predetermined criterion, in particular, after predetermined time intervals and/or each time word data is received during operating time of the speech recognizer.

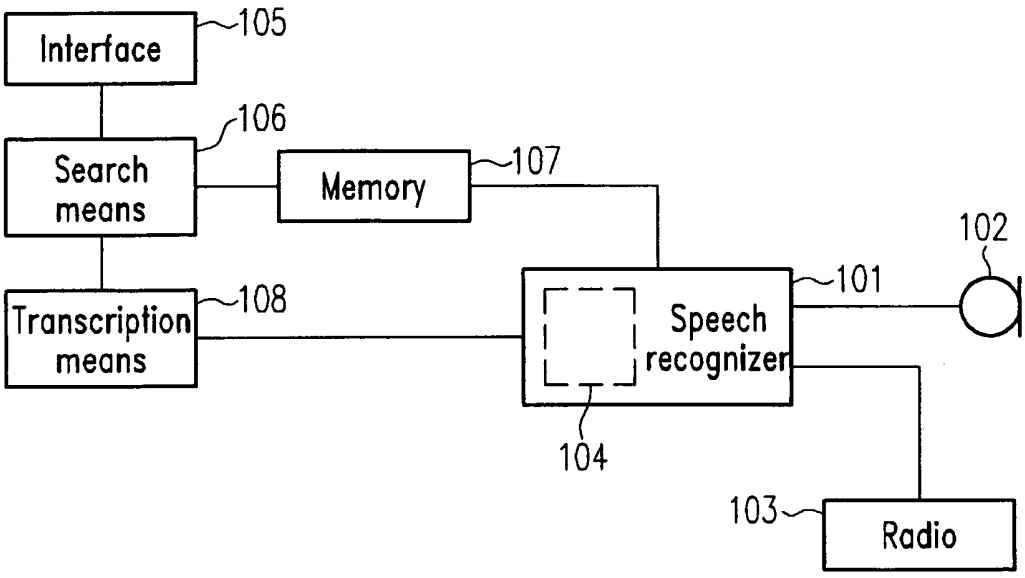

Interface
105
Search means
106
Memory
107
Transcription means
108
Speech recognizer
101
102
104
103
Radio

Fig.1

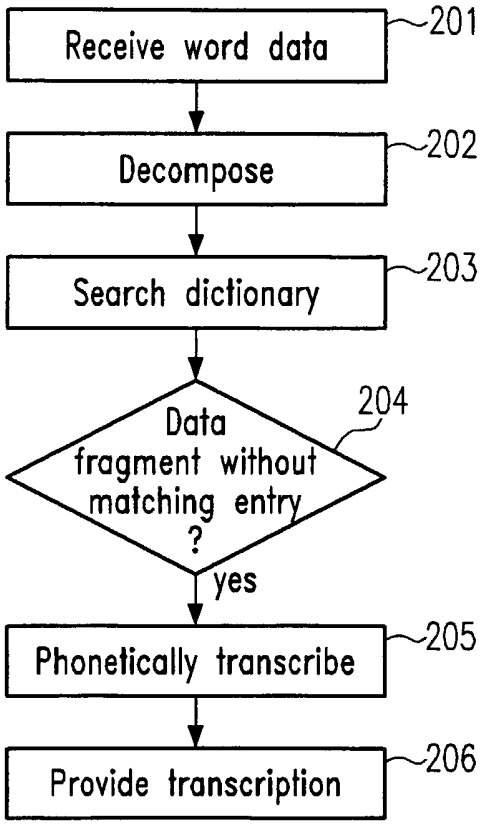

Receive word data
201
Decompose
202
Search dictionary
203
Data fragment without matching entry ?
204
yes
Phonetically transcribe
205
Provide transcription
206

Fig.2

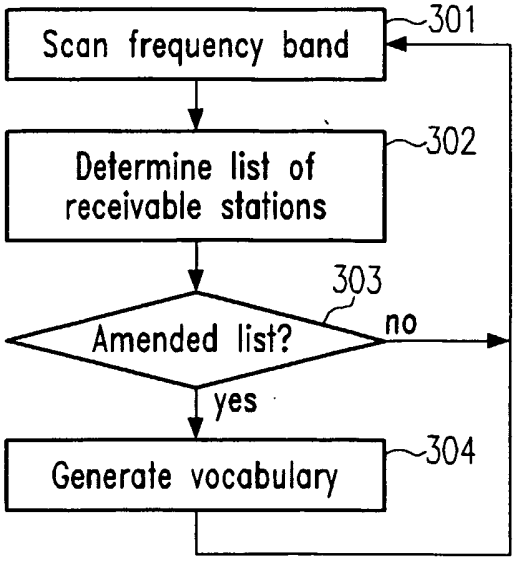

Scan frequency band
301
Determine list of receivable stations
302
Amended list?
303
no
yes
Generate vocabulary
304

Fig.3

SWR 1	92,6 MHz
SWR 2	89,2 MHz
SWR 3	97,4 MHz
SWR 4 HN	99,5 MHz
Radio Energy	100,7MHz
⋮	⋮

Fig.4

SWR	Südwestrundfunk
Phonetical transcription	*Phonetical transcription*
Energy	
Phonetical transcription1	*Phonetical transcription2*
HN	Heilbronn
Phonetical transcription	*Phonetical transcription*
⋮	

Fig.5

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP2005/005568

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L13/08 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)
IPC 7 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 practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 0 735 736 A (AT & T CORP) 2 October 1996 (1996-10-02) abstract	1,13,14
Y	HAIN H: "Ein hybrider Ansatz zur Graphem-Phonem-Konvertierung unter Verwendung eines Lexikons und eines neuronalen Netzes" ELEKTRONISCHE SPRACHSIGNALVERARBEITUNG KONFERENZ, XX, XX, 4 September 2000 (2000-09-04), pages 160-167, XP002223265 abstract paragraph '0002!	1,13,14

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 October 2005	27/10/2005

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

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP2005/005568

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LAMMENS J M G ED - LAVER J ET AL LAVER J ET AL: "A LEXICON-BASED GRAPHEME-TO-PHONEME CONVERSION SYSTEM" EUROPEAN CONFERENCE ON SPEECH TECHNOLOGY. EDINBURGH, SEPTEMBER 1987, EDINBURGH, CEP CONSULTANTS, GB, vol. VOL. 1 CONF. 1, 1 September 1987 (1987-09-01), pages 281-284, XP000010730 page 282, line 43 - page 283, line 5 * p.283 "The Algorigthm * * p.283 "Abbreviations may be converted or expanded" * * p.284 line 13-16 *	1,13,14
A	DIVAY M ED - ACOUSTICAL SOCIETY OF JAPAN: "A WRITTEN TEXT PROCESSING EXPERT SYSTEM FOR TEXT TO PHONEME CONVERSION" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (ICSLP). KOBE, NOV. 18 -22, 1990, TOKYO, ASJ, JP, vol. VOL. 2, 18 November 1990 (1990-11-18), pages 853-856, XP000506904 abstract paragraph '0III!	2-4, 15-17
A	TRANCOSO I ET AL: "ON THE PRONUNCIATION MODE OF ACRONYMS IN SEVERAL EUROPEAN LANGUAGES" 5TH EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY. EUROSPEECH '97. RHODES, GREECE, SEPT. 22 - 25, 1997, EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY. (EUROSPEECH), GRENOBLE : ESCA, FR, vol. VOL. 2 OF 5, 22 September 1997 (1997-09-22), pages 573-576, XP001003959 paragraph '0004!	5,6,10, 18,19,23
A	EP 1 220 200 A (SIEMENS AG) 3 July 2002 (2002-07-03) column 4, line 14 - line 22 column 5, line 1 - line 15	1,13,14

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP2005/005568

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0735736 A	02-10-1996	DE 69633883 D1	30-12-2004
		ES 2233954 T3	16-06-2005
		JP 3561076 B2	02-09-2004
		JP 8320696 A	03-12-1996
		US 5724481 A	03-03-1998
		US 5822727 A	13-10-1998
EP 1220200 A	03-07-2002	DE 50003855 D1	30-10-2003
		ES 2208212 T3	16-06-2004