

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
1 April 2004 (01.04.2004)

PCT

(10) International Publication Number
WO 2004/027752 A1

(51) International Patent Classification⁷: **G10L 13/00**,
13/08

(21) International Application Number:
PCT/US2003/029137

(22) International Filing Date:
16 September 2003 (16.09.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/251,354 20 September 2002 (20.09.2002) US

(71) Applicant: **MOTOROLA, INC., A CORPORATION
OF THE STATE OF DELAWARE** [US/US]; 1303 East
Algonquin Road, Schaumburg, IL 60196 (US).

(72) Inventors: **MA, Changxue**; 4929 Lichfield Drive, Bar-
rington, IL 60010 (US). **RANDOLPH, Mark**; 21922
Barkley Court, Kildeer, IL 60047 (US).

(74) Agents: **NICHOLS, Daniel K.** et al.; MOTOROLA,
INC., Intellectual Property Department, 1303 East Algo-
nquin Road, Schaumburg, IL 60196 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, 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, KP, KR,
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT,
RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR,
TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

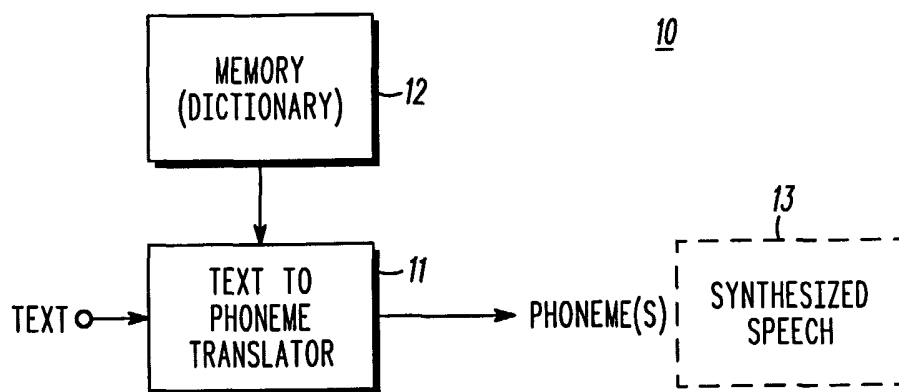
(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,
SE, SI, SK, TR), OAPI patent (BF, BJ, 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 "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND APPARATUS TO FACILITATE CORRELATING SYMBOLS TO SOUNDS



(57) Abstract: A dictionary (12) is comprised of a dendroid hierarchy of branches and nodes, wherein each node represents no more than one symbol (which symbol is to be converted to a corresponding sound) and wherein each such symbol as is represented at a given node has only one corresponding sound associated with that symbol at that node. In addition, many of the branches include a plurality of nodes representing a string of the symbols in a particular sequence. The dictionary (12) is used to translate an input comprising a given integral sequence of the symbols into a corresponding integral sequence of sounds (11). This permits both method and apparatus to convert, for example, text to representative phonemes (11). Such phonemes can be used, amongst other purposes, to support synthesized speech production (13).

WO 2004/027752 A1

METHOD AND APPARATUS TO FACILITATE CORRELATING SYMBOLS TO SOUNDS

Technical Field

- 5 This invention relates generally to the correlation of symbols to sounds and more particularly to the conversion of text to phonemes.

Background

Prior art approaches exist to convert text into corresponding sounds.

- 10 Such techniques permit, for example, the conversion of text into audible synthesized speech. Many such approaches use phonemes that are units of a phonetic system of the relevant spoken language and that are usually perceived to be single distinct sounds in the spoken language. Using phonemes in this way in fact constitutes a relatively effective and accurate mechanism to achieve telling results. Unfortunately,
- 15 however, prior art techniques do not always reliably select the correct phonemes.

- Part of the problem stems from the fact that, in many spoken languages that have a corresponding symbolic alphabet, one or more of the symbols have more than one proper pronunciation. As a result, some symbols have more than one potentially appropriate phoneme (or set of phonemes) associated therewith. Various prior art
- 20 approaches have been suggested to attempt mitigating the effect of this circumstance. Unfortunately, these solutions generally tend to be computationally intensive and/or require a considerable amount of memory. This tends to render such solutions inappropriate for use in resource-limited platforms (such as, for example, cellular

telephones) where computational capacity itself and/or electric power can be considerably constrained.

For example, one prior art approach (known in at least some circles as “N-gram analysis”) uses a combination of probability analysis and grammatical context to weight a corresponding conclusion regarding pronunciation of a given word. To illustrate, the word “read” can be enunciated in English in either of two ways depending upon the grammatical context. By storing the rules regarding such context and by examining other words around the word “read” in view of those rules, one can potentially deduce a correct pronunciation for a given instance of the word. Again, however, such an approach often requires at least a significant quantity of memory as well as a fairly elaborate development and manipulation of contextual rules.

Many prior art approaches also fall short in view of another common occurrence; the need to pronounce a proper name or other word that is not in the dictionary of the process. To ameliorate, at least to some extent, this problem, the prior art suggests permitting a user to train the process by introducing the word along with its pronunciation. This approach, however, can be time consuming, tedious, confusing to the user, and again highly consumptive of memory and computational capacity.

20 Brief Description of the Drawings

The above needs are at least partially met through provision of the method and apparatus to facilitate correlating symbols to sounds described in the following detailed description, particularly when studied in conjunction with the drawings, wherein:

FIG. 1 comprises a block diagram view of a text to speech platform as configured in accordance with an embodiment of the invention;

FIG. 2 comprises a general flow diagram as configured in accordance with an embodiment of the invention;

5 FIG. 3 comprises a detailed flow diagram as configured in accordance with an embodiment of the invention;

FIG. 4 comprises a schematic view of an illustrative portion of a hierarchically organized dictionary as configured in accordance with an embodiment of the invention;

10 FIG. 5 comprises a lattice view that illustrates selection of a given branch within the hierarchically organized dictionary as configured in accordance with an embodiment of the invention; and

FIG. 6 comprises a detailed portion of a flow diagram as configured in accordance with another embodiment of the invention.

15 Skilled artisans will appreciate that elements in the figures are illustrated for simplicity and clarity and have not necessarily been drawn to scale. For example, the dimensions of some of the elements in the figures may be exaggerated relative to other elements to help to improve understanding of various embodiments of the present invention. Also, common but well-understood elements that are useful or
20 necessary in a commercially feasible embodiment are typically not depicted in order to facilitate a less obstructed view of these various embodiments of the present invention.

Detailed Description

Generally speaking, pursuant to these various embodiments, a symbol-to-sound translator (such as a text to phoneme translator) utilizes a dictionary comprising a dendroid hierarchy of branches and nodes, wherein each node represents no more than one of the symbols and wherein each such symbol as is represented at a node has
5 only one corresponding sound associated with that symbol at that node, and where each branch includes a plurality of nodes representing a string of the symbols in a particular sequence. In a preferred embodiment, at least some of the symbols comprise alphanumeric textual characters such as letters. If desired, a combination of symbols can be used to represent a single sound (such as the combination of letters
10 "ch" that can be used in the English language to represent a single phoneme sound). Also in a preferred embodiment, at least some of the sounds can be comprised of phonemes. If desired, the strings of symbols as represented by the branches can represent entire words in the corresponding spoken language. In a preferred embodiment, however, such strings can also accommodate incomplete words such as,
15 but not limited to, grammatical prefixes, suffixes, stems, and/or morphemes.

In a preferred embodiment, at least some of the nodes have a probability indicator correlated therewith. This indicator reflects how frequently the corresponding sound associated with the symbol at that node has been previously selected for use when translating an input that included the symbol at that node. If
20 desired, such probability indicators can be recalculated and revised dynamically on a substantially continuous basis. In a alternative embodiment, a probability indicator located in one portion of a branch can be used to temporarily impact the probability indicator as associated with a node located elsewhere in that same branch. For example, the probability of use indicator for a given node can be modified as a

function of at least one probability of use indicator for a lower hierarchical node on a shared branch. In a preferred embodiment, this modification comprises temporarily replacing the probability indicator at the given node with the probability indicator for the node located lower in the dictionary dendroid hierarchy.

5 Referring now to the drawings, and in particular FIG. 1, a symbol-to-sound platform 10 will typically include a text to phoneme translator 11 having a memory 12 either operably coupled thereto or internally contained therein. The memory 12, in addition to such other content (such as programming instructions and/or other data as may be used by the text to phoneme translator 11) as may be stored therein, includes a
10 dictionary. In this embodiment, the dictionary comprises a dendroid hierarchy of branches and nodes, wherein each node represents no more than one symbol and wherein each such symbol as is represented at a node has only one corresponding sound associated with that symbol at that node. In general, each branch includes a plurality of nodes. The plurality of nodes represents a string (or plurality of strings) of
15 the symbols in a particular sequence (in a preferred embodiment, these strings include a variety of complete words as well as grammatical prefixes, suffixes, stems, and morphemes). Such strings can correspond to more than one written/spoken language if desired, but in a preferred embodiment are largely directed to only a single language per dictionary (and, of course, multiple dictionaries as correspond to
20 different language can be simultaneously stored in the memory 12). At least some of the symbols will appear repeatedly at different nodes with different corresponding sounds. Additional description regarding such a dictionary appears below. In general, the symbol-to-sound platform 10 comprises a programmable platform such as a microprocessor, microcontroller, programmable gate array, digital signal processor, or

the like (though if desired, a less flexible platform architecture could be used where appropriate to a given application).

The text to phoneme translator 11 has one or more inputs to receive symbols. In this embodiment, at least some of the symbols comprise alphanumeric textual characters and in particular comprise combined alphanumeric textual characters such as a series of words comprising a plurality of sentences. Such text can be sourced to support a variety of different purposes. For example, the text may correspond to a word processing document, a webpage, a calculation or enquiry result, or any other text source that the user wishes, for whatever reason, to hear audibly enunciated.

10 In this embodiment, the text to phoneme translator 11 produces sounds comprised of phonemes (where phonemes are understood to each comprise units of a phonetic system of spoken language that are perceived to be single distinct sounds in the spoken language). Typically, a given integral sequence of symbols introduced at the input will yield a corresponding integral sequence of sounds at the output. For example, a first integral sequence of letters that comprise a single word will yield a corresponding integral sequence of phonemes that represent an audible utterance of that particular word. If desired, such phoneme information can be used to facilitate, for example, the synthesization of speech 13. Phoneme information can be used for other purposes as well, however, and these teachings are applicable for use in such alternative applications as well.

Such a symbols-to-sounds platform 10 can be a standalone platform or can be comprised as a part of some other device or mechanism, including but not limited to computers, personal digital assistants, telephones (including wireless and cordless

telephones), and various consumer, retail, commercial, and industrial object interfaces.

Referring now to FIG. 2, in the various embodiments presented herein, in general a dictionary (or dictionaries) having a dendroid hierarchy is provided 21 and
5 used to translate 22 symbol input (such as text input) into corresponding sounds (such as phonemes). As described above, a memory 12 can serve to provide such a dictionary and a text to phoneme translator 11 can serve to so translate symbols into corresponding sounds.

Referring now to FIG. 3, the symbol to sound process will be described in
10 more detail. As already noted, the platform 10 receives input comprising one or more symbols (such as alphanumeric text). For example, the input can comprise the alphanumeric expression "gone," which includes four letters combined to form a single word in the English language. Each of these letters has a corresponding sound (which "sound" can include silence, of course) and, at least in the English language,
15 will typically have a number of corresponding sounds. These embodiments serve to facilitate the correct choosing of such sounds to achieve a proper pronunciation of the word itself as represented by the appropriate phonemes. Such integral symbol groups are parsed 32 to separate the individual characters. For example, the word "gone" would be parsed into the individual letters "g," "o," "n," and "e." The platform then
20 identifies 33 appropriate corresponding nodes in the dictionary.

Referring momentarily to FIG. 4, this concept of nodes and the overall dendroid hierarchy of the dictionary will be described in more detail. Each node in the dictionary hierarchy includes a single symbol and a single corresponding sound. There can be multiple nodes, however, that share a common symbol. Such nodes will

also typically have differing sounds. For example, there can be a plurality of nodes 41 that each include the letter “g” 42 and 43. The first node 42, however, can have a corresponding sound S^1 for the symbol “g” such as the sound of “g” in the English word “give,” while a second node 43 has a corresponding sound S^2 such as the sound of “g” in the English word “gin.”

Each such node may then couple via a branch to one or more other nodes. For example, the first “g” node 42 noted above can couple to a number of other nodes 44 including a node 45 that includes the letter “o” and the corresponding sound S^3 of “o” as occurs in the English word “song” (the other nodes 44 can include the same letter “o” and/or other letters entirely – for example, one node might include the letter “i” as part of the string “give”). In a similar fashion, this secondary node with the letter “o” 45 can itself branch to another hierarchical level 46 to represent yet additional symbols such as a node for the letter “n” (with corresponding sound S^4 for the letter “n” pronounced as in the English word “con”) (and as part of a hierarchical branch that includes the string “gone”) and a node for the letter “i” (with corresponding sound S^5 for the letter “i” pronounced as in the English word “stopping”) (and as part of a hierarchical branch that includes the string “going”).

So configured, it should be evident that many words and word parts are readily represented as strings of such nodes and that duplicate letter/sound entries are avoided to some extent by the dendroid hierarchical structure described. As a result, a dictionary composed in such a way can represent a relatively large quantity of textual input (and corresponding phoneme content) in a relatively small amount of memory.

In addition, a probability indicator (or indicators) can be also provided at some (or all) nodes to provide an indication of how frequently the corresponding sound

associated with the symbol at that node has been selected for use when translating an input that included the symbol at that node. In particular, such an indicator can represent how many times the corresponding sound for the symbol at a given node has been selected as compared to identical symbols having different corresponding sounds at other nodes at the same hierarchical level as the given node. Such probabilities can be calculated apriori and included as a static component of the dictionary. In a preferred embodiment, however, the probability indicators are dynamic and change in value with experience and use of the dictionary. The probabilities can all begin at an equal level of probability (or can be initially offset as desired) and can then be recalculated as desired to update the probability indicators.

For example, and with continued reference to FIG. 4, the first "g" node 42 described above can have a probability indicator C^1 associated therewith (such as ".6") and the second "g" node 43 can have a probability indicator C^2 associated therewith (such as ".4"). Such values would indicate that the sound S^1 for the first "g" node 42 has been used more often than the sound S^2 for the second "g" node 43.

So configured, and referring now back again to FIG. 3, the platform 10 can next determine 34 the probability of use as corresponds to each previously identified node by accessing the probability indicator for each such node. With such information, the platform 10 can then select 35 a most likely hierarchical branch for the text input now being processed. There are a variety of ways that such a selection can be effected. In a preferred embodiment, and referring momentarily to FIG. 5, the candidate nodes and their corresponding probability indicators can be conceptually represented as a lattice. A "most likely" path through the lattice will result in identifying a particular hierarchical branch for the given text. To illustrate this

concept, a lattice presents the probability indicators for each candidate node for the individual letters of the text "gone." For purposes of this example, a first candidate sound at a first node 51 for the letter "g" has a probability indicator of ".4." This probability indicator is less than the probability indicator of ".6" as exists for a second candidate sound at a second node 52 for the letter "g." As a result, the second candidate sound as associated with the probability indicator of ".6" is selected. In a similar fashion, the highest probability indicator for each group of candidate nodes for each letter is in turn selected until a complete branch has been identified for the text.

Returning again to FIG. 3, the platform 10 then selects 36 the corresponding sounds for each node of the resulting hierarchical branch. These corresponding sounds are, in this example, the phonemes that constitute the output of the process.

In a process where the probability indicators are dynamically altered through use, the probability indicators can now be updated 37 to reflect this most recent use of the dictionary to select a particular sequence of phonemes to represent a given text input.

In a preferred embodiment, and referring now to FIG. 6, subsequent to determining 34 the probabilities of use of the various candidate nodes and prior to selecting 35 the most likely hierarchical branch, the platform 10 can modify 61 one or more of the probability of use indicators. In particular, a higher probability node that is lower on the hierarchical scale can be used to more significantly weight a lower probability node that is higher on the hierarchical scale. To illustrate, the probability indicator for a given node that is higher than the probability indicator for another node that shares the same hierarchical branch as the given node and that is higher on that branch than the given node can have its probability indicator substituted for the

probability indicator of the hierarchically lower node. (In another embodiment, if desired, the probability indicator of the hierarchically higher node can be modified in other ways, such as by taking an average of the two probability indicators.)

Viewed in a more rigorous light, consider that the probability $P(\beta_1, \beta_2, \dots, \beta_n | \alpha_1, \alpha_2, \dots, \alpha_m)$ indicates the likelihood for a given phone sequence $\beta_1, \beta_2, \dots, \beta_n$ as a whole being generated from a given text string $\alpha_1, \alpha_2, \dots, \alpha_m$. Pursuant to the above process, pronunciations for all possible sub-strings of the input are retrieved from the dendroid hierarchical dictionary and this probability is calculated as the sum of the probabilities for all possible phonetic realizations for the input sub-strings. For a given input word $\omega = \alpha_1, \alpha_2, \dots, \alpha_m$, let $\omega_i^k(j) = \alpha_i \dots \alpha_{j-1} \alpha_j \alpha_{j+1} \dots \alpha_k$ denote the sub-string of word ω beginning in position i with letter α_i , ending in position k with letter α_k , and having a focus letter α_j . In other words, $\alpha_i \dots \alpha_{j-1}$ and $\alpha_{j+1} \dots \alpha_k$ denote α_j 's left and right context respectively. Paths $\tau_{ik}^{(j)}$ in the hierarchical context tree are a set of letter-to-sound translations of $\omega_i^k(j)$ found by search the dictionary tree, where $k \geq j$. Basically, as the search extends letter by letter from left to right, the context tree grows. If no letter match is found the context tree stops growing.

For each input word string, the platform 10 searches the dictionary repeatedly until all possible pronunciations of a given input sub-string are found. In other words, the search starts at each node of the dictionary tree until each of the nodes has been used as a starting node. In this way, the occurrence of each path $\tau_{ik}^{(j)}$ will be accumulated.

In many cases the dictionary will not include the whole text string. Nevertheless, in most cases, at least some partial segments of the text string will typically be found in the dictionary. A variable context length can therefore be used in this method as the sum of the probabilities for all the relevant input letter sequences.

In this way, the occurrence of each path $\tau_{ik}^{(j)}$ will be accumulated. To illustrate, let $N(\alpha_i, \alpha_{i+1}, \dots, \alpha_k)$ represent the counts for string segment $\alpha_i, \alpha_{i+1}, \dots, \alpha_k$ and let $M(\beta_i^l, \beta_{i+1}^l, \dots, \beta_k^l)$ represent the counts for its l th transcription. The probability for transcription $\beta_i^l, \beta_{i+1}^l, \dots, \beta_k^l$ can therefore be estimated as:

$$P(\beta_i^l, \beta_{i+1}^l, \dots, \beta_k^l | \alpha_i, \alpha_{i+1}, \dots, \alpha_k) = \frac{M(\beta_i^l, \beta_{i+1}^l, \dots, \beta_k^l)}{N(\alpha_i, \alpha_{i+1}, \dots, \alpha_k)}$$

s

These probabilities comprise the probability indicators that are recorded at the leaf nodes of the context trees as described earlier. It should be noted that for each node in the context tree, there can be more than one probability associated with it, because the node can have more than one child node. With the first Viterbi pass, the probabilities on the leaf nodes propagate upwards and retain the maximum probability value for each node.

In effect, for each new word, the process chooses a letter as the focus and uses maximum possible context around the focused letter. The process then uses this word segment as a key to traverse the dendroid hierarchy of the dictionary. During this traversal, sub-trees are generated. These sub-trees contain all possible context segments ranging from a minimum length to maximum length. To start the tree traversal at any node of the dictionary tree, the counts $M(\beta_i^l, \beta_{i+1}^l, \dots, \beta_k^l)$ and

$N(\alpha_i, \alpha_{i+1}, \dots, \alpha_k)$ of how an orthographic segment is transformed into a pronunciation are accumulated.

After building the sub-tree, the probabilities of symbol to phoneme mapping at each level of the sub-tree are estimated. The probabilities at the leaf node of the sub-
5 tree are then propagated upwardly with respect to the hierarchical structure of the tree. In a preferred embodiment, when the probability of mapping on a child node is larger than that of the parent, then the probability indicator for the parent node is replaced with that of the child node.

All the paths $\tau_{ik}^{(j)}$ in the sub-trees are translated into a lattice representation for
10 generating N-best baseform transcriptions with a Viterbi search. To consider the edge effects where a given cut point could lose important context information, a window function that centers on the focused grapheme letters can be used to weigh down the contribution of the probabilities near both ends of the text string. Since the probabilities are estimated for each grapheme in the text with all possible context
15 lengths, the probability of each grapheme is a mixture of all windowed segment probabilities. Penalties can also be added to adjust the weight for segments of different length. In general, a shorter context will be accorded a higher penalty because long contexts offer more disambiguation than shorter ones.

It should be observed that the focused letters whose phonemes are searched for
20 can consist of a consonant string or a vowel string. This means that the process can obtain the corresponding phonemes without breaking the consonant or vowel strings. This can aid in avoiding a lot of unnecessary and misleading conversions. Also, each occurrence of the context segment is counted. Therefore the longest segment and the

most frequent one play a dominant role in determining the letter-to-sound conversion. Further, the dictionary can be built up recursively so that it covers the data where basic rules can be learned. These basic rules should predict a significant part of the big dictionary accurately.

5 So configured, the resultant dictionary and corresponding process are relatively well suited to facilitate various symbol-to-sound activities in a way that potentially requires less memory than prior approaches. In addition, the described platform and processes are well suited in particular to support the pronunciation of words that are not actually included in the dictionary for whatever reason, thereby
10 meeting a significant existing need.

 Those skilled in the art will recognize that a wide variety of modifications, alterations, and combinations can be made with respect to the above described embodiments without departing from the spirit and scope of the invention, and that such modifications, alterations, and combinations are to be viewed as being within the
15 ambit of the inventive concept.

We claim:

1. A method of correlating symbols with sounds, wherein at least some of the symbols can correspond to a plurality of sounds and at least some of the sounds can correspond to a plurality of symbols, comprising:
 - 5 providing a dictionary comprising a dendroid hierarchy of branches and nodes, wherein each node represents no more than one of the symbols and wherein each such symbol as is represented at a node has only one corresponding sound associated with that symbol at that node, and each branch includes a plurality of nodes representing a string of the symbols in a particular sequence; and
 - 10 automatically using the dictionary to translate an input comprising a given integral sequence of the symbols into a corresponding integral sequence of sounds.

2. The method of claim 1 wherein providing a dictionary comprising a dendroid hierarchy of branches and nodes further includes correlating a probability indicator with each such symbol as is represented at a node to provide an indication of how frequently the corresponding sound associated with the symbol at that node has been
5 selected for use when translating an input that included the symbol at that node, and wherein automatically using the dictionary to translate an input comprising a given integral sequence of the symbols into a corresponding integral sequence of sounds includes using at least one of the probability indicators to translate the input into the corresponding integral sequence of sounds.

10

3. The method of claim 1 wherein automatically using the dictionary to translate an input comprising a given integral sequence of the symbols into a corresponding integral sequence of sounds includes:
receiving a first plurality of symbols that, together and in the given integral sequence,
15 represents an expression in a spoken language; and
accessing the dendroid hierarchy of branches and nodes to identify nodes having corresponding symbols that correlate to the individual symbols that comprise the first plurality of symbols to form a plurality of candidate corresponding sounds, and wherein providing a dictionary comprising a dendroid hierarchy of branches and
20 nodes further includes correlating a probability of use indicator with each such symbol as is represented at a node to provide an indication of how frequently the corresponding sound associated with the symbol at that node has been selected for use when translating an input that included the symbol at that node.

4. The method of claim 3 wherein automatically using the dictionary to translate an input comprising a given integral sequence of the symbols into a corresponding integral sequence of sounds further includes using the probability of usage indicator as is associated with at least some of the symbols that correspond to the nodes to
5 select a particular corresponding sound from amongst the plurality of candidate corresponding sounds, and wherein correlating the probability of use indicator with each such symbol as is represented at a node includes calculating the probability of use indicator for each such symbol as a function of how many times the corresponding sound for the symbol at a given node has been selected as compared to
10 identical symbols having different corresponding sounds at other nodes at the same hierarchical level as the given node.

5. The method of claim 4 wherein correlating the probability of use indicator with each such symbol as is represented at a node further includes modifying the
15 probability of use indicator for a given node as a function of at least one probability of use indicator for a lower hierarchical node located elsewhere on a branch that includes the given node.

6. The method of claim 5 wherein modifying the probability of use indicator for a
20 given node as a function of at least one probability of use indicator for a lower hierarchical node located on a branch that includes the given node includes at least temporarily replacing the probability of use indicator for a given node with the probability of use indicator for the lower hierarchical node.

7. An apparatus comprising:

a memory having a dictionary stored therein, the dictionary comprising a dendroid hierarchy of branches and nodes, wherein each node represents no more than one symbol and wherein each such symbol as is represented at a node has only one

5 corresponding sound associated with that symbol at that node, and each branch includes a plurality of nodes representing a string of the symbols in a particular sequence, wherein at least some of the symbols appear repeatedly at different nodes with different corresponding sounds; and

a text-to-phoneme translator operably coupled to the memory, the text-to-phoneme

10 translator includes translation means for converting text into phonemes as a function, at least in part, of the contents of the dictionary.

8. The apparatus of claim 7 wherein the dictionary further includes a probability of use indicator for at least some of the nodes as corresponds to the represented symbol

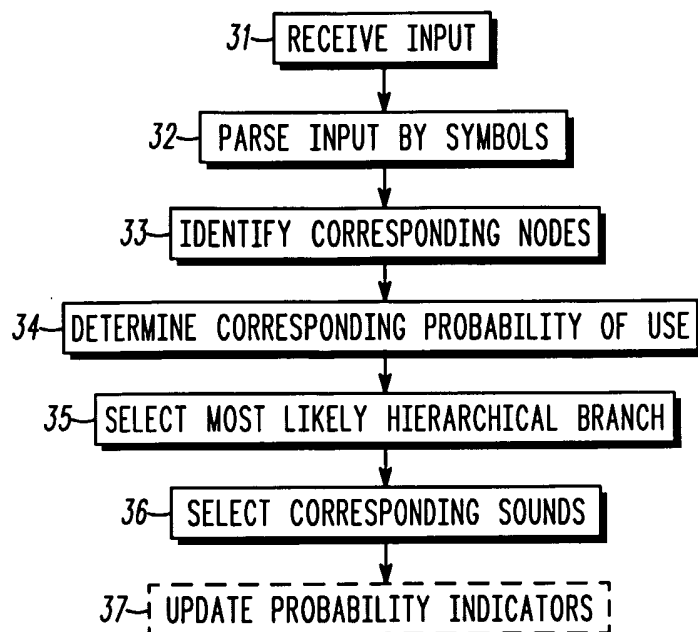
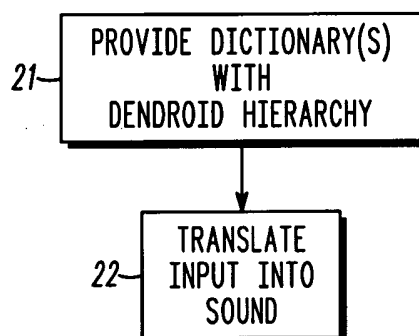
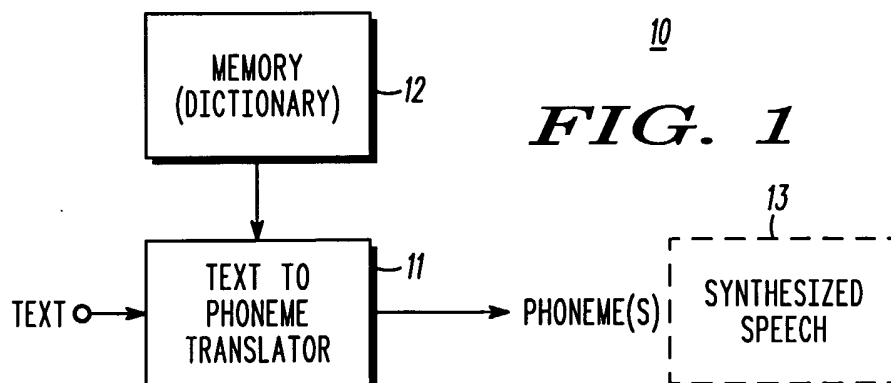
15 and the corresponding sound associated therewith, wherein the translation means further converts text into phonemes as a function, at least in part, of the probability of use indicators.

9. The apparatus of claim 8 wherein the translation means further at least temporarily

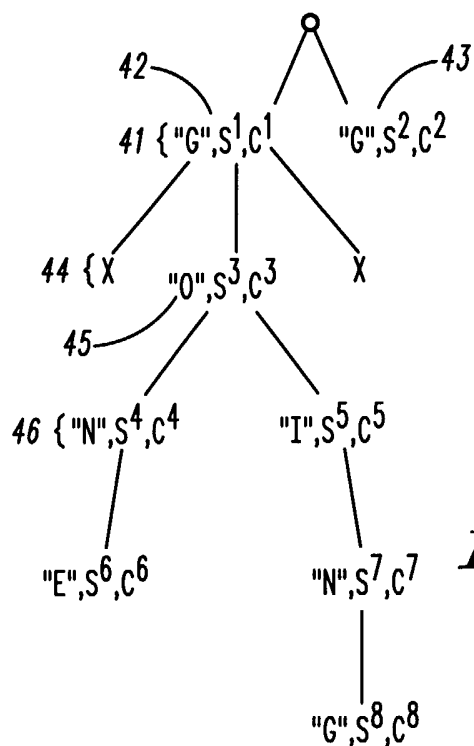
20 alters at least one probability of use indicator to facilitate selection of a given corresponding sound to use when translating text into the phonemes.

10. The apparatus of claim 9 wherein the translation means alters the at least one probability of use indicator as a function, at least in part, of other probability of use indicators as are retained in the dictionary.

1/2



2/2



40
FIG. 4

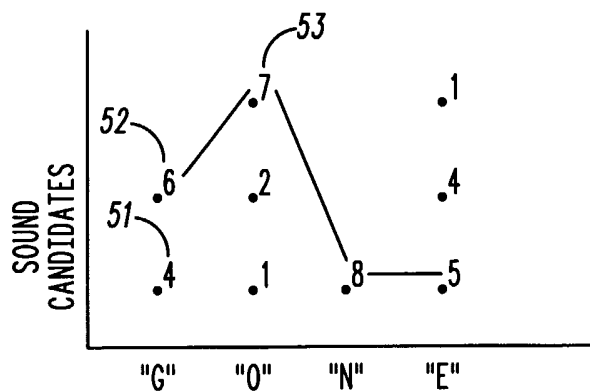


FIG. 5

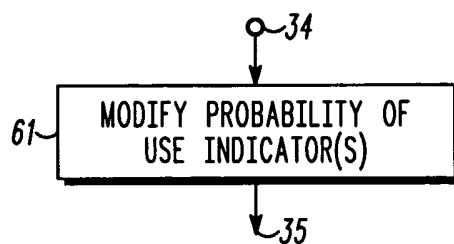


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/29137

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G10L 13/00; 13/08

US CL : 704/258, 266

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)

U.S. : 704/258, 260, 263, 266, 269

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)

Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6,016,471 A (KUHN et al.) 18 January 2000 (18.01.2000), column 2, lines 23 to 29;	1, 7
---	column 2, lines 43 to 61; column 2, lines 61 to 65; column 2, line 66 to column 3, line 30;	-----
Y	column 4, lines 55 to 60; Figures 1 to 3.	2-6, 8-10
Y	US 6,363,342 B2 (SHAW et al.) 26 March 2002 (26.03.2002), column 3, line 29 to column 4, line 40; column 5, lines 3 to 6; column 8, lines 19 to 34; column 9, lines 19 to 29; Figures 3 to 5.	2-6, 8-10
A	US 5,682,501 A (SHARMAN) 28 October 1997 (28.10.1997), column 8, lines 28 to 47; column 10, lines 29 to 47; Figure 3.	1-10
A	US 6,347,295 B1 (VITALE et al.) 12 February 2002 (12.02.2002), column 2, line 60 to column 3, line 39; column 8, lines 6 to 49; column 11, line 5 to column 12, line 23; Figure 5.	1-10



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 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

27 January 2004 (27.01.2004)

Date of mailing of the international search report

01 MAR 2004

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Facsimile No. (703)305-3230

Authorized officer

Richemond Dorvil

Telephone No. (703) 305-4700

INTERNATIONAL SEARCH REPORT

PCT/US03/29137

Continuation of B. FIELDS SEARCHED Item 3:

EAST search terms: node and probability and speech adj synthes\$; and homony\$; and counts; and occur\$ adj frequenc\$; and frequen\$ with use