

Description

This invention relates to a speech recognition system, and to a method of constructing templates for use in recognising speech.

In known speech recognition systems, words uttered are subjected to known pattern recognition techniques, and, if correspondence with a known word is found, suitable coded signals are generated identifying the word. Correspondence is generally determined by generating signals or "scores" indicating the degree of similarity with stored patterns corresponding to known words, the word having the best score being deemed to be the word uttered.

This type of technique suffers from the disadvantage that the success rate of identifying words is largely dependent upon the quality of the stored patterns (templates) of the words. The usual way of building template sets of words is to generate these from a large number of spoken versions of the words; that is to say by asking a large number of people to speak the words, and to analyse the resultant speech patterns. The words are input by reading aloud a list of the words.

Another method of constructing templates for use in recognising speech was related in the IEEE 1986 National Aerospace and Electronics Conference, see the paper by T. P. Barry and D. T. Williamson at pages 699 to 802 of part 3 of the proceedings. The method uses a software programme entitled "Software Methodology for Automated Recognition Training SMART". The SMART process is capable of forming representative speech templates of words which are spoken by a user. The user is prompted by the system to speak the words in two different contexts:

The first context is to simply say the words as if being read from a list;

The second context involves repeating different sentences into which the specific word is embedded.

Several separate templates of the word are created for both contexts, the templates then being combined by the software programme to form a single composite template for that word.

The system is capable of reprompting, at any point and in either context, if the user provides an inadequate response to a prompt.

Alternatively, each of the input people is asked to repeat the words of a spoken list. In either case, the words are uttered in a "forced" mode rather than in a "free" mode (that is to say when the words are used in their proper context). It has been found that many speakers pronounce words slightly differently in these two modes, so by storing templates of words spoken in the "forced" mode, and subsequently comparing these templates with the speech patterns of the same words spoken in the "free" mode, can lead to a reduced success rate in word identification.

According to a first aspect the present invention provides a method of constructing, for a list of words, templates for use in speech recognition, the method comprising: the steps of analyzing speech signals from a plurality of users to construct at least one template for each word from said list which is input by each user, and constructing from said templates a set of representative templates for each of said words, wherein during said constructing step each user inputs all the words of a given list of words in a "forced" mode, either by reading the words of said list or by repeating the words of the list which have first been spoken by a synthesizer, and then inputs at least some of the words of the list in a "free" mode by answering a series of questions spoken by said synthesizer, said questions being designed to elicit as one-word answers from the user words from the list. Advantageously, each word input in the "forced" mode is compared with all the reference templates of the words in the list, these reference templates being stored in a first template store, and in the analyzing step for each instance of a word spoken in the "forced" mode a template created for that word instance is flagged (labelled) and then stored in the second template store.

In a preferred embodiment, the analyzing step compares for each instance of a word spoken in the "free" mode a template created for that instance with all the templates stored in the first and second template stores, the template for that instance being identified and labelled, the labelled template being stored in the second template store.

Advantageously, all the templates of a given number of users are stored in a third template store prior to the constructing step, and the templates identified during the constructing step are stored in a fourth template store.

Advantageously the constructing step includes the step of comparing each template in the third template store with all the templates of the same word, and with all the templates of any other word associated with said first-mentioned word in a predetermined manner, the constructing step being such that the templates of a given word stored in the fourth template store are representative of a maximum number of the templates of that word stored in the third template store and unrepresentative of any other template stored in the template store.

According to a second aspect the present invention provides a method of generating a set of speech templates as in claim 6.

According to a third aspect the present invention provides a speaker-independent isolated-word speech recognition system as in claim 7.

One embodiment of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:-

Figure 1 is a block diagram of a speech recogni-

tion apparatus; and

Figure 2 is a block diagram of sifting apparatus.

Referring to the drawings, Figure 1 shows a speech recognition apparatus in which speech from a user is received by a microphone 1 connected to a speech recogniser 2. The recogniser 2 compares received words with the contents of a reference template store 3 which contains representations of a repertoire of words which it is desired to recognise.

Any of a number of conventional recognition algorithms may be used for this comparison, and these will not therefore be discussed in detail. By way of example, the "VOTAN" recogniser card produced by Votan Inc. for use with an IBM PC microcomputer might be employed.

The recogniser 2 compares a received word with each of the stored representations, and produces for each a similarity signal or "score" which indicates the closeness of fit between the two. Normally, the word whose stored representation has the best score is the one "recognised" and a corresponding coded signal is passed via a line 4 to a control unit 5, which could for example be the aforementioned IBM computer, for onward transmission or to initiate further action. For example, the control unit 5 can respond by generating a message back to the user via a speech synthesiser 7 and a loudspeaker 8. The synthesiser 7 has a parameter store 9 to enable it to generate appropriate words to be sent to the user. A further line 6 is provided between the recogniser 2 and the control unit 5, this further line being used (in a manner described below) when two or more stored representations have the same or similar scores.

All the words spoken by the user are stored in an audio signal store 10, which may be, for example, a tape recorder. After the recogniser 2 has compared each received word with each of the stored representations, and has "recognised" a word, the speech pattern (or template) of that word is stored in a template store 11. Each template stored in the template store 11 is flagged to indicate which word it is.

In use, a user first speaks a list of words, either by reading from a list or by repeating a list of words generated by the synthesiser 7. During this "forced" mode, each spoken word is compared with all the stored representations in the reference template store 3, and the template initially flagged to correspond to the best match is stored in the template store 11, as described above. After all the words in the list have been spoken in the "forced" mode, the user is asked to respond to a series of questions generated by the control unit 5 and "spoken" by the speech synthesiser 7. The control unit 5 generates these questions via suitable software, and the questions are designed to elicit one-word answers from the user, these words being from the list. In this way, the user speaks words from the list more in context, that is to say in the "free" mode. Whenever the user gives an answer, that word

is compared, by the recogniser 2, with all the stored representations in the reference template store 3 and with all the templates in the template store 11. It then stores the template of that word, suitably flagged to correspond to the best match, in the template store 11.

The entire process is then repeated with a large number of other users. In each case, each spoken word (in either the "forced" mode or the "free" mode) is compared with all the templates in both the reference template store 3 and the template store 11.

When a sufficiently large number of templates have been stored (usually 1000 or more), these can be analysed to provide an improved initial template set. This is accomplished as described below with reference to Figure 2, which shows schematically a template store 20 which contains all the templates of, say, 30 male users and 30 female users. The templates of each group of 60 templates of the same word are then analysed in a sifter 21 to determine which of the templates are the best. To do this, each template is compared with all the templates of the same word, and with all the templates of any other word in the same node. (A node is a group of words which are linked together as being probable answers to a given question - for example "yes" and "no" belong to one node, and a series of numbers belongs to another node.) The sifter 21 uses a conventional algorithm which picks out six male templates and six female templates as the best for each word. This is accomplished by testing each template against all other templates of the same node group in such a manner that the twelve chosen templates recognise a maximum number of correct words, and no incorrect words. In this way, the sifting process reduces the number of templates from 2×30 to 2×6 . It should be noted that it is not possible to use a single template to recognise a given word, in view of the fact that different speakers give significantly different speech patterns depending upon their accents and physiology. It has been found that six templates for each set to gives satisfactory recognition accuracy.

Obviously, where there are more than 60 original users templates stored in the template store 11, it will be necessary to modify the sifting process. One way of doing this would be to have a plurality of sets of 60 (2×30) templates, each of which is reduced to sets of 12 (2×6) templates, and then to make further sets of 60 (2×30) templates from the sets of 12 (2×6) templates for further runs through the sifter 21. Alternatively, once a reduced set of 12 (2×6) templates has been produced, all the other templates can be compared with this reduced set to see whether the word concerned is correctly recognised. The templates of unrecognised words can then be used to modify the original reduced set by a further sifting process.

The entire sifting process is carried out for each of the words stored, so that finally each word to be

recognised has a set of 2 x 6 templates. These templates have an increased success rate in subsequent word identification processes, because the original templates include templates of words spoken in the "free" mode.

If, during the initial comparison process (in either the "forced" mode or the "free" mode), the recogniser 2 gives two scored representations the same score or similar scores, the result is ambiguous, and a signal indicating this is passed, via the line 6, along with codes for both words, via the line 4, to the control unit 5 (see Figure 1). The control unit 5 responds by generating a message back to the user via the speech synthesiser 7 and the loudspeaker 8.

This message has the form of "Did you say X" where X is the word whose representation stored in the reference template store 3 gave rise to the better score (or, if the two scores were identical, one of the two selected at random); and awaits a reply.

If the user replies "Yes" (or "No") this is recognised by the recogniser 2, and signalled to the control unit 5 which, in the event of a Yes proceeds as if X had been identified originally. In the event of a "No", a further message is issued via the synthesiser 7, viz "Did you say Y". Again the user response is analysed and if Y is confirmed, recognition is deemed complete. If the user again replies "No", the control unit 5 then initiates generation of a request for repetition (although in principle of course a third choice could be offered). This method of resolving ambiguity is described in greater detail in GB Patent Specification no. 2192746.

A particularly convenient way of arranging for users to provide a required group of words in both the "forced" and "free" mode is by using an automated telephone competition. For example, one competition may relate to Telephone Banking, in which case each caller would be asked to repeat a list of words generated by the synthesiser 7 from the parameter store 9. This list may comprise the words "one", "two", "three", "four", "five", "six", "seven", "eight", "nine", "zero", "oh", "nought", "current", "deposit", "savings", "Visa", "Access", "statement", "mini-statement", "transfer", "chequebook", "balance", "help", "pounds", "pence", "full", "yes", "no", and "stop".

As mentioned above, these words would be repeated by each caller so as to generate a series of the "forced" mode templates. The control unit 5 would then instruct the synthesiser 7 to put a series of questions to the caller, these questions being designed to elicit one-word answers from the list. In order to ensure a suitably high number of entrants (and hence a sufficiently large number of template-generating users), it is advisable for the competition to offer a prize or prizes. The audio store 10 is useful in weeding out spurious inputs, for example words deliberately incorrectly input.

Claims

1. A method of constructing, for a list of words, templates for use in speech recognition, the method comprising the steps of:
 - analysing speech signals from a plurality of users to construct at least one template for each word from said list which is input by each user, and
 - constructing from said templates a set of representative templates for each of said words, characterised in that during said constructing step each user inputs all the words of a given list of words in a "forced" mode, either by reading the words of said list or by repeating the words of the list which have first been spoken by a synthesizer, and then inputs at least some of the words of the list in a "free" mode by answering a series of questions spoken by the said synthesizer, said questions being designed to elicit as one-word answers from the user words from the list.
2. A method as in claim 1, wherein:
 - each word input in the "forced" mode is compared with all the reference templates of the words in the list, these reference templates being stored in a first template store,
 - in the analysing step for each instance of a word spoken in the "forced" mode a template created for that word instance is identified by comparison with each of the stored reference templates, and the template created for that word instance is labelled and then stored in a second template store.
3. A method as in claim 2, wherein the analysing step compares for each instance of a word spoken in the "free" mode a template created for that instance with all the templates stored in the first and second template stores, the template for that instance being identified and labelled, the labelled template being stored in the second template store.
4. A method as in claim 3, wherein:
 - all the templates of a given number of users are stored in a third template store prior to the constructing step, and
 - the templates identified during the constructing step are stored in a fourth template store.
5. A method as in claim 4, wherein:
 - the constructing step includes the step of comparing each template in the third template store with all the templates of the same word, and with all the templates of any other word associated with said first-mentioned word in a predeter-

recognised has a set of 2 x 6 templates. These templates have an increased success rate in subsequent word identification processes, because the original templates include templates of words spoken in the "free" mode.

If, during the initial comparison process (in either the "forced" mode or the "free" mode), the recogniser 2 gives two scored representations the same score or similar scores, the result is ambiguous, and a signal indicating this is passed, via the line 6, along with codes for both words, via the line 4, to the control unit 5 (see Figure 1). The control unit 5 responds by generating a message back to the user via the speech synthesiser 7 and the loudspeaker 8.

This message has the form of "Did you say X" where X is the word whose representation stored in the reference template store 3 gave rise to the better score (or, if the two scores were identical, one of the two selected at random); and awaits a reply.

If the user replies "Yes" (or "No") this is recognised by the recogniser 2, and signalled to the control unit 5 which, in the event of a Yes proceeds as if X had been identified originally. In the event of a "No", a further message is issued via the synthesiser 7, viz "Did you say Y". Again the user response is analysed and if Y is confirmed, recognition is deemed complete. If the user again replies "No", the control unit 5 then initiates generation of a request for repetition (although in principle of course a third choice could be offered). This method of resolving ambiguity is described in greater detail in GB Patent Specification no. 2192746.

A particularly convenient way of arranging for users to provide a required group of words in both the "forced" and "free" mode is by using an automated telephone competition. For example, one competition may relate to Telephone Banking, in which case each caller would be asked to repeat a list of words generated by the synthesiser 7 from the parameter store 9. This list may comprise the words "one", "two", "three", "four", "five", "six", "seven", "eight", "nine", "zero", "oh", "nought", "current", "deposit", "savings", "Visa", "Access", "statement", "mini-statement", "transfer", "chequebook", "balance", "help", "pounds", "pence", "full", "yes", "no", and "stop".

As mentioned above, these words would be repeated by each caller so as to generate a series of the "forced" mode templates. The control unit 5 would then instruct the synthesiser 7 to put a series of questions to the caller, these questions being designed to elicit one-word answers from the list. In order to ensure a suitably high number of entrants (and hence a sufficiently large number of template-generating users), it is advisable for the competition to offer a prize or prizes. The audio store 10 is useful in weeding out spurious inputs, for example words deliberately incorrectly input.

Claims

1. A method of constructing, for a list of words, templates for use in speech recognition, the method comprising the steps of:

analysing speech signals from a plurality of users to construct at least one template for each word from said list which is input by each user, and

constructing from said templates a set of representative templates for each of said words, characterised in that during said constructing step each user inputs all the words of a given list of words in a "forced" mode, either by reading the words of said list or by repeating the words of the list which have first been spoken by a synthesizer, and then inputs at least some of the words of the list in a "free" mode by answering a series of questions spoken by the said synthesizer, said questions being designed to elicit as one-word answers from the user words from the list.

2. A method as in claim 1, wherein:

each word input in the "forced" mode is compared with all the reference templates of the words in the list, these reference templates being stored in a first template store,

in the analysing step for each instance of a word spoken in the "forced" mode a template created for that word instance is identified by comparison with each of the stored reference templates, and the template created for that word instance is labelled and then stored in a second template store.

3. A method as in claim 2, wherein the analysing step compares for each instance of a word spoken in the "free" mode a template created for that instance with all the templates stored in the first and second template stores, the template for that instance being identified and labelled, the labelled template being stored in the second template store.

4. A method as in claim 3, wherein:

all the templates of a given number of users are stored in a third template store prior to the constructing step, and

the templates identified during the constructing step are stored in a fourth template store.

5. A method as in claim 4, wherein:

the constructing step includes the step of comparing each template in the third template store with all the templates of the same word, and with all the templates of any other word associated with said first-mentioned word in a predeter-



mined manner, the constructing step being such that the templates of a given word stored in the fourth template store are representative of a maximum number of the templates of that word stored in the third template store and unrepresentative of any other templated stored in the third template store.

6. A method of generating a set of speech templates, which set is to be used in a subsequent speech recognition process, the method comprising the steps of: for each of a multiplicity of words
 - i) generating a plurality of first speech templates in a "forced" mode where a human speaker is prompted directly to speak a given word, and
 - ii) generating a plurality of second speech templates in a "free" mode where a human speaker is induced to speak the same given word but without direct request for same, for example, by being asked a question to which said word is calculated to form the answer; and
 - iii) selecting a final set of templates from said pluralities of first and second speech templates, which final set is subsequently to be used in a speech recognition process, the final set containing no composite templates formed from said first and/or said second speech templates.
7. A speaker-independent isolated-word speech recognition system having a multi-word vocabulary and incorporating a stored set of templates, the stored set of templates including for each word of said vocabulary a plurality of templates, characterised in that said stored set of templates includes templates produced from "free" mode utterances but no composite templates formed from "free" mode and/or "forced" mode utterances, the stored set of templates being constructed according to the method of any one of claims 1 to 6.

Patentansprüche

1. Verfahren zur Erzeugung von Schablonen für eine Liste von Wörtern zur Verwendung bei der Spracherkennung, mit den Schritten:
Analysieren von Sprachsignalen von mehreren Benutzern, um zumindest eine Schablone für jedes Wort aus der Liste zu erzeugen, das durch jeden Benutzer eingegeben wurde, und
Erzeugen aus diesen Schablonen eines Satzes von repräsentativen Schablonen für jedes der Worte,
dadurch gekennzeichnet, daß

im Erzeugungsschritt jeder Benutzer alle Worte einer gegebenen Wortliste in einem "Zwangs"-Modus eingibt, indem er entweder die Wörter der Liste liest oder indem er die Wörter der Liste wiederholt, die vorher durch einen Synthesizer ausgesprochen wurden, und dann zumindest einige der Wörter der Liste in einem "freien" Modus eingibt, indem er eine Reihe von durch den Synthesizer ausgesprochenen Fragen beantwortet, wobei die Fragen so entworfen sind, daß sie vom Benutzer die Wörter der Liste als Einwort-Antworten hervorrufen.

2. Verfahren nach Anspruch 1, bei dem:
jedes im "Zwangs"-Modus eingegebene Wort mit allen Bezugsschablonen der Wörter in der Liste verglichen wird, wobei diese Bezugsschablonen in einem ersten Schablonenspeicher gespeichert sind,
im Analyseschritt für jedes im "Zwangs"-Modus ausgesprochene Einzelwort eine für dieses Einzelwort erzeugte Schablone identifiziert wird, indem sie mit jeder der gespeicherten Vergleichsschablonen verglichen wird, wobei die für das Einzelwort erzeugte Schablone markiert und dann in einem zweiten Schablonenspeicher gespeichert wird.
3. Verfahren nach Anspruch 2, bei dem im Analyseschritt für jedes im "freien" Modus ausgesprochene Einzelwort eine für dieses Einzelwort erzeugte Schablone mit allen im ersten und zweiten Schablonenspeicher gespeicherten Schablonen verglichen wird, wobei die Schablone für dieses Einzelwort identifiziert und markiert und die markierte Schablone im zweiten Schablonenspeicher gespeichert wird.
4. Verfahren nach Anspruch 3, bei dem:
alle Schablonen einer gegebenen Anzahl von Benutzern vor dem Erzeugungsschritt in einem dritten Schablonenspeicher gespeichert werden, und
die im Erzeugungsschritt identifizierten Schablonen in einem vierten Schablonenspeicher gespeichert werden.
5. Verfahren nach Anspruch 4, bei dem:
der Erzeugungsschritt den Schritt des Vergleichens jeder Schablone im dritten Schablonenspeicher mit allen Schablonen des gleichen Worts und mit allen Schablonen beliebiger anderer Worte, die mit dem erstgenannten Wort in vorbestimmter Weise verbunden sind, umfaßt, wobei der Erzeugungsschritt so ausgelegt ist, daß die im vierten Schablonenspeicher gespeicherten Schablonen eines gegebenen Worts repräsentativ für eine Maximalzahl von im dritten



Schablonenspeicher gespeicherten Schablonen dieses Wort sind und nicht repräsentativ für beliebige andere Schablonen, die im dritten Schablonenspeicher gespeichert sind.

6. Verfahren zur Erzeugung eines Satzes von Sprachschablonen, der in einem nachfolgenden Spracherkennungsprozeß verwendet wird, wobei das Verfahren für jedes einer Mehrzahl von Worten umfaßt:

i) Erzeugen mehrerer erster Sprachschablonen in einem "Zwangs"-Modus, in dem ein menschlicher Sprecher unmittelbar aufgefordert wird, ein gegebenes Wort auszusprechen, und

ii) Erzeugen mehrerer zweiter Sprachschablonen in einem "freien" Modus, in dem ein menschlicher Sprecher dazu veranlaßt wird, das gleiche gegebene Wort auszusprechen, ohne daß er direkt darum gebeten worden wäre, beispielsweise indem er eine Frage gefragt wird, auf die zur Bildung der Antwort das Wort berechnet wird,

gekennzeichnet durch den Schritt

iii) Auswählen eines endgültigen Satzes von Schablonen aus den ersten und zweiten Sprachschablonen, wobei der endgültige Satz nachfolgend in einem Spracherkennungsvorgang verwendet wird, wobei der endgültige Satz keine aus den ersten und/oder zweiten Sprachschablonen zusammengesetzten Schablonen umfaßt.

7. Sprecherunabhängiges Spracherkennungssystem für isolierte Worte mit einem Mehrwortvokabular, wobei das System einen gespeicherten Satz von Schablonen aufweist, wobei der gespeicherte Schablonensatz für jedes Wort des Vokabulars mehrere Schablonen aufweist, dadurch gekennzeichnet, daß

der gespeicherte Satz von Schablonen Schablonen enthält, die aus Aussprachen im "freien" Modus hergeleitet wurden, jedoch keine zusammengesetzten Schablonen, die aus Aussprachen im "freien" Modus und/oder im "Zwangs"-Modus gebildet werden würden, wobei der gespeicherte Satz von Schablonen gemäß dem Verfahren nach einem der Ansprüche 1 bis 6 aufgebaut ist.

Revendications

1. Procédé de construction, pour une liste de mots, de modèles destinés à servir dans la reconnaissance de la parole, le procédé comprenant les étapes consistant à :

analyser des signaux de fréquence vocale provenant d'une multiplicité d'utilisateurs pour construire un modèle au moins pour chaque mot,

issu de ladite liste, qui est entré par chaque utilisateur, et

construire à partir desdits modèles un jeu de modèles représentatif pour chacun desdits mots, caractérisé en ce que, pendant ladite étape de construction, chaque utilisateur entre la totalité des mots d'une liste donnée de mots en un mode "forcé", soit en lisant les mots de ladite liste, soit en répétant les mots de la liste qui ont été d'abord prononcés par un synthétiseur puis entre au moins certains des mots de la liste en mode "libre" par réponse à une suite de questions prononcées par ledit synthétiseur, lesdites questions étant conçues pour choisir des réponses d'un mot à partir des mots prononcés par l'utilisateur issus de la liste.

2. Procédé selon la revendication 1, dans lequel : chaque mot entré en mode "forcé" est comparé à tous les modèles de référence des mots figurant dans la liste, ces modèles de référence étant mémorisés dans une première mémoire de modèles,

au cours de l'étape d'analyse pour chaque formulation d'un mot prononcé en mode "forcé", un modèle créé pour cette formulation de mot est identifié en comparaison avec chacun des modèles de référence mémorisés, le modèle créé pour cette formulation de mot étant repéré puis mémorisé dans une seconde mémoire de modèles.

3. Procédé selon la revendication 2, dans lequel l'étape d'analyse compare pour chaque cas d'un mot prononcé en mode "libre" un modèle créé pour ce cas à tous les modèles mémorisés dans les première et seconde mémoires de modèles, le modèle pour ce cas étant identifié et repéré, le modèle repéré étant mémorisé dans la seconde mémoire de modèles.

4. Procédé selon la revendication 3, dans lequel : tous les modèles d'un nombre donné d'utilisateurs sont mémorisés dans une troisième mémoire de modèles préalablement à l'étape de construction, et

les modèles identifiés pendant l'étape de construction sont mémorisés dans une quatrième mémoire de modèles.

5. Procédé selon la revendication 4, dans lequel : l'étape de construction comprend l'étape

consistant à comparer chaque modèle présent dans la troisième mémoire de modèles à tous les modèles du même mot et à tous les modèles de tout autre mot quelconque associé audit premier mot mentionné d'une manière prédéterminée, l'étape de construction étant telle que les modèles d'un mot donné mémorisés dans la quatrième



mémoire de modèles sont représentatifs d'un nombre maximal des modèles de ce mot mémorisés dans la troisième mémoire de modèles et non représentatifs de tout autre modèle mémorisé dans la troisième mémoire de modèles.

5

6. Procédé de génération d'un jeu de modèles de parole, lequel jeu est destiné à être utilisé dans un processus de reconnaissance de parole ultérieur, le procédé comprenant les étapes consistant à :
- i) pour chacun d'une multiplicité de mots
 - i) générer une multiplicité de premiers modèles de parole en mode "forcé", dans lequel un locuteur humain est invité directement à prononcer un mot donné, et
 - ii) générer une multiplicité de seconds modèles de parole en un mode "libre" dans lequel un locuteur humain est induit à prononcer le même mot donné, sans cependant d'invite directe pour se faire, par exemple, en étant invité à répondre à une question à laquelle ledit mot est calculé pour constituer la réponse, et caractérisé par l'étape consistant à :
 - iii) sélectionner un jeu final de modèles à partir desdites multiplicités de premiers et seconds modèles de parole, lequel jeu final est destiné à être ultérieurement utilisé dans un processus de reconnaissance de parole, le jeu final ne contenant aucun modèle composite formé à partir desdits premier et second modèles de parole.
7. Système de reconnaissance de la parole de mots isolés indépendants du locuteur comportant un vocabulaire de mots multiples et incorporant un jeu mémorisé de modèles, le jeu mémorisé de modèles comprenant pour chaque mot dudit vocabulaire une multiplicité de modèles, caractérisé en ce que ledit jeu mémorisé de modèles comprend des modèles produits à partir d'énonciations en mode "libre" mais aucun modèle composite formé à partir d'énonciations en mode "libre" et/ou en mode "forcé", le jeu mémorisé de modèles étant construit conformément au procédé de l'une quelconque des revendications 1 à 6.

10

15

20

25

30

35

40

45

50

55

7



* C L M S *