

Description

The present invention relates to voice-operated services, for example services (such as home banking) which may be offered on a telephone network. In particular, the invention concerns allocation and recognition of user identifiers which a user must enter in order to identify himself or herself to the apparatus operating the service.

From Weston Technical Papers, 30 October - 2 November 1984 "Practical Experience with a prototype speaker verification system" D.E. Crabbs et al it is known to reject a user if too many 'rejection' decisions are produced. However a valid user may be continuously rejected because the user is unable to enunciate the required words clearly enough.

The present invention in its various aspects is defined in the claims.

Note that the word "template" is used to signify data representing the sound of a word, with which a comparison can be made; it is not intended to imply any particular type of speech recognition.

The apparatus shown in the figure forms part of the control equipment of a telephone banking service. A telephone line 1 is connected to a speech recogniser 2 and a speech synthesizer 3. A controller 4 (e.g. a microprocessor system or other digital computer) can transfer commands to the synthesizer 3 to cause desired words to be uttered and receives data output from the recogniser 2.

The recogniser, of conventional design, compares words received over the telephone line with a set of templates and produces at its data output

- (i) a code W1 indicating the template which most closely matches
- (ii) a code W2 indicating the next closest template
- (iii) scores D1, D2 indicating the goodness of fit between the word and those two templates.

In order to use the service, an initial log-on procedure requires the user to identify himself or herself by uttering a plurality of words, preferably a series of digits, though other words could of course be used. A five digit log-on number is hereafter assumed. The resulting codes W1 are received by the controller 4 and thus identify the user. Further steps (with which we are not here concerned) then follow; for example the checking of a personal identity number (PIN) - for which of course speaker dependent recognition (i.e. using templates specific to the particular user) may now be employed. Speaker independent recognition of course must necessarily be used to recognise the log-on number.

The aspect of the system with which we are here particularly concerned is the allocation of the log-on number for a new user, the object being to avoid digits which represent a high probability of mis-recognition during subsequent log-on procedures.

The user is asked to utter the words one, two, three, four, five, six, seven, eight and nine. These voice samples along with any other desired words, may be recorded for the formation of templates for speaker-dependent recognition. For present purposes, however, they are passed to the speech recogniser 2, which produces, for each word, the codes W1 and W2, and scores D1, D2 as indicated in table 1 below

Table 1

	one	two	three		nine
W1	1	3	3		
W2	4	2	4		
D1	18	34	21		
D2	25	29	42		

The codes shown for W1, W2 are simply the actual numbers which are deemed by the recogniser to have been recognised. The scores are numbers, on an arbitrary scale, indicating the extent of deviation of the received word from the template; thus a low score indicates a good match whilst a high score indicates a poor match. The scores given are assumed to be actual results for a particular user. Looking at the example given in Table 1, it can be seen that the words, "one" and "three" have been correctly recognised, whilst the word "two" has been mistaken for "three".

The controller 4, having received this data applies certain criteria to ascertain which digits (if any) offer such poor reliability of recognition that they should be excluded from the log-on number. For the example user of Table 1, the word two is rejected since mis-recognition has actually occurred. However the remaining digits are also examined. For reliable recognition, what is important is not that the score obtained should be low, but that the score obtained with the "correct" template is significantly better than with the second choice template. Thus in the example the word "three" has scores of 21 and 42 and will therefore

be more reliably recognised than "one" where the scores, though lower, are similar and mis-recognition may occur. Thus, in this embodiment, the criterion applied is that $(D2-D1)$ is greater than some threshold T .

The controller 4, computes the difference $(D2-D1)$ for each digit, compares the differences with the threshold T and rejects any digits which do not meet the criterion. It then allocates to the user a five-digit log-on number none of whose digits is one which has failed to meet the criterion. Clearly, if more acceptable digits are available than are needed, it is preferable to use those with the larger differences. This allocation can be in accordance with any convenient procedure - eg, selection of the next unused number which does not contain a rejected digit.

The threshold T can be selected on the basis of experience at a level falling between one at which digits are rarely rejected and one at which many digits are consistently rejected. With a practical selection for T it may occur, with users having particularly poor diction, that eight or even nine digits are rejected. Therefore the controller may reduce the threshold T for that particular user down to a level such that a desired minimum number of digits meet the criterion - for example, it may be decided that a five-digit log-on number should contain at least three mutually different digits.

Two modifications to this procedure will now be described. In the first, it is assumed that the controller, when operating a log-on procedure, is capable of resolving ambiguities in recognition - ie a situation where scores $D1$ and $D2$ are the same or differ by a small amount (say 5) - by offering the first choice back to the user (via the speech synthesizer 3) and asking for confirmation, and, if this is not forthcoming, offering the second choice. (Such arrangements are the object of our pending UK patent application no. 2192746A).

In order to allow in extreme cases the use of digits which are incorrectly recognised (but where the second choice is correct) during the allocation procedure, the criterion is modified as follows:

- (i) if neither the first nor second best score is incorrect, is given a zero rating;
- (ii) if the first is correct, the rating is defined as $(D2-D1)+5$ (ie occupying a rating range from 5 upwards);
- (iii) if the second is correct, the rating is defined as $5-(D2-D1)$: a negative result is deemed to be zero (ie those cases where the difference is less than 5 occupy the rating range from 1 to 4).

The rating is then compared with the threshold value, in the same manner as previously.

In the second modification, cognisance is taken of the fact that recognition may be improved by using different templates for male and female speakers. In this case, during the allocation process, speech recognition scores are obtained using both sets of templates. The controller determines which set gives the best results (ie which set gives the largest $(D1-D2)$ values, or which set gives measures which are in general larger than does the other) and carries out the remainder of the allocation process using the results from that set only.

If the first digit of the allocated number is chosen to indicate whether the user is male or female, for example by allocating 1, 2 or 3 in the former case and 4, 5 or 7 in the latter, then the recogniser can be arranged during the log-on process to use, for the first digit, male templates for 1, 2 and and female templates for 4, 5 and 7. The set of templates to be used for recognition of the remaining few digits is then determined by the recognition result from the first digit.

Claims

1. A method of allocating to a user of a voice operated service a user identifier comprising a plurality of words from a set of words, comprising:
 - (i) requesting the user to utter the words of the set;
 - (ii) comparing each uttered word with a set of speech recognition templates which correspond to the words of the set to obtain a measure of their similarity;
 - (iii) determining which of the words of the set meet a criterion as to the extent to which the measure of similarity of the uttered word to the template corresponding to the requested word exceeds the measure of similarity of the uttered word to that one of the other templates with which the uttered word has the greatest similarity; and
 - (iv) selecting a sequence of words using only words of the set which meet the criterion.
2. A method according to Claim 1 in which the criterion is that the difference of the two measures of similarity exceeds a threshold value.
3. A method according to Claim 1 or 2 in which:
 - (a) each uttered word is compared with two sets of templates;
 - (b) the measures of similarity obtained with the two sets are compared to identify the set which gives measures which are in general larger than does the other set; and

be more reliably recognised than "one" where the scores, though lower, are similar and mis-recognition may occur. Thus, in this embodiment, the criterion applied is that $(D2-D1)$ is greater than some threshold T .

The controller 4, computes the difference $(D2-D1)$ for each digit, compares the differences with the threshold T and rejects any digits which do not meet the criterion. It then allocates to the user a five-digit log-on number none of whose digits is one which has failed to meet the criterion. Clearly, if more acceptable digits are available than are needed, it is preferable to use those with the larger differences. This allocation can be in accordance with any convenient procedure - eg, selection of the next unused number which does not contain a rejected digit.

The threshold T can be selected on the basis of experience at a level falling between one at which digits are rarely rejected and one at which many digits are consistently rejected. With a practical selection for T it may occur, with users having particularly poor diction, that eight or even nine digits are rejected. Therefore the controller may reduce the threshold T for that particular user down to a level such that a desired minimum number of digits meet the criterion - for example, it may be decided that a five-digit log-on number should contain at least three mutually different digits.

Two modifications to this procedure will now be described. In the first, it is assumed that the controller, when operating a log-on procedure, is capable of resolving ambiguities in recognition - ie a situation where scores $D1$ and $D2$ are the same or differ by a small amount (say 5) - by offering the first choice back to the user (via the speech synthesizer 3) and asking for confirmation, and, if this is not forthcoming, offering the second choice. (Such arrangements are the object of our pending UK patent application no. 2192746A).

In order to allow in extreme cases the use of digits which are incorrectly recognised (but where the second choice is correct) during the allocation procedure, the criterion is modified as follows:

- (i) if neither the first nor second best score is incorrect, is given a zero rating;
- (ii) if the first is correct, the rating is defined as $(D2-D1) + 5$ (ie occupying a rating range from 5 upwards);
- (iii) if the second is correct, the rating is defined as $5-(D2-D1)$: a negative result is deemed to be zero (ie those cases where the difference is less than 5 occupy the rating range from 1 to 4).

The rating is then compared with the threshold value, in the same manner as previously.

In the second modification, cognisance is taken of the fact that recognition may be improved by using different templates for male and female speakers. In this case, during the allocation process, speech recognition scores are obtained using both sets of templates. The controller determines which set gives the best results (ie which set gives the largest $(D1-D2)$ values, or which set gives measures which are in general larger than does the other) and carries out the remainder of the allocation process using the results from that set only.

If the first digit of the allocated number is chosen to indicate whether the user is male or female, for example by allocating 1, 2 or 3 in the former case and 4, 5 or 7 in the latter, then the recogniser can be arranged during the log-on process to use, for the first digit, male templates for 1, 2 and and female templates for 4, 5 and 7. The set of templates to be used for recognition of the remaining few digits is then determined by the recognition result from the first digit.

Claims

1. A method of allocating to a user of a voice operated service a user identifier comprising a plurality of words from a set of words, comprising:
 - (i) requesting the user to utter the words of the set;
 - (ii) comparing each uttered word with a set of speech recognition templates which correspond to the words of the set to obtain a measure of their similarity;
 - (iii) determining which of the words of the set meet a criterion as to the extent to which the measure of similarity of the uttered word to the template corresponding to the requested word exceeds the measure of similarity of the uttered word to that one of the other templates with which the uttered word has the greatest similarity; and
 - (iv) selecting a sequence of words using only words of the set which meet the criterion.
2. A method according to Claim 1 in which the criterion is that the difference of the two measures of similarity exceeds a threshold value.
3. A method according to Claim 1 or 2 in which:
 - (a) each uttered word is compared with two sets of templates;
 - (b) the measures of similarity obtained with the two sets are compared to identify the set which gives measures which are in general larger than does the other set; and



(c) the determination as to which words meet the criterion is carried out using the identified set.

4. A method according to Claim 1 or 2 in which:

(a) each uttered word is compared with two sets of templates;

(b) the set for which the differences (D1-D2) between the measures of similarities of each uttered word to the corresponding requested word, and to that one of the other templates of that set with which the uttered word has the greater similarity, are in general largest is identified; and

(c) the determination as to which words meet the criterion is carried out using the identified set.

5. A method according to Claim 3 or 4 in which two sets of templates correspond to male and female speech respectively.

6. A method according to Claim 3, 4 or 5 in which, in the selection of a sequence of words, the first of the words is taken from one of two subsets of the set of words, in dependence on which of the two sets of templates is identified as giving the larger measures.

7. An apparatus for allocating a user identifier comprising a speech recogniser and control means connected to receive output data from the recogniser and arranged in operation to perform the method of any one of Claims 1 to 6.

8. An apparatus according to Claim 7 for operating an interactive voice-operated service accessible via a telephone line.

9. An apparatus according to Claim 8 in which the voice operated service is a home banking service.

Patentansprüche

1. Verfahren zum Zuteilen eines Benutzerkennzeichners an einen Benutzer eines sprachbetriebenen Dienstes, wobei der Kennzeichner eine Vielzahl von Wörtern aus einem Wortsatz umfaßt, und wobei das Verfahren aufweist:

(i) Auffordern des Benutzers, Wörter des Wortsatzes auszusprechen;

(ii) Vergleichen jedes ausgesprochenen Wortes mit einem Satz von Spracherkennungsmustern, die den Wörtern des Satzes entsprechen, um ein Maß ihrer Ähnlichkeit zu erhalten;

(iii) Bestimmen, welches der Wörter des Satzes ein Kriterium in bezug auf den Grad erfüllt, bis zu dem das Maß der Ähnlichkeit des ausgesprochenen Wortes, im Vergleich zum Muster entsprechend dem geforderten Wort, das Maß der Ähnlichkeit des ausgesprochenen Wortes im Vergleich zu demjenigen der anderen Muster überschreitet, mit denen das ausgesprochene Wort die größte Ähnlichkeit besitzt; und

(iv) Wählen einer Wortfolge unter Benutzung nur solcher Worte des Satzes, die das Kriterium erfüllen.

2. Verfahren nach Anspruch 1, bei dem das Kriterium darin besteht, daß der Unterschied der beiden Ähnlichkeitsmaße einen Schwellenwert überschreitet.

3. Verfahren nach Anspruch 1 oder 2, bei dem:

(a) jedes ausgesprochene Wort mit zwei Sätzen von Mustern verglichen wird;

(b) die mit den zwei Sätzen erhaltenen Ähnlichkeitsmaße verglichen werden, um denjenigen Satz zu identifizieren, der Maße liefert, die im allgemeinen größer als diejenigen sind, die der andere Satz liefert; und

(c) die Bestimmung um herauszufinden, welche Wörter das Kriterium erfüllen, unter Benutzung des identifizierten Satzes durchgeführt wird.

4. Verfahren nach Anspruch 1 oder 2, bei dem:

(a) jedes ausgesprochene Wort mit zwei Sätzen von Mustern verglichen wird;

(b) der Satz, bei dem die Unterschiede (D1-D2) zwischen den Ähnlichkeitsmaßen jedes ausgesprochenen Wortes im Vergleich zum entsprechenden geforderten Wort und zu demjenigen der anderen Muster des Satzes, bei dem das ausgesprochene Wort die größte Ähnlichkeit besitzt, allgemein am größten sind, identifiziert wird; und



(c) die Bestimmung, um herauszufinden, welche Wörter das Kriterium erfüllen, unter Benutzung des identifizierten Satzes durchgeführt wird.

- 5 5. Verfahren nach Anspruch 3 oder 4, bei dem zwei Mustersätze dem männlichen bzw. dem weiblichen Sprachausdruck entsprechen.
- 10 6. Verfahren nach den Ansprüchen 3, 4 oder 5, bei dem bei der Wahl einer Folge von Wörtern das erste der Wörter aus einer von zwei Untergruppen des Wortsatzes entnommen wird, wobei in Abhängigkeit von diesem Wort derjeniger der beiden Mustersätze identifiziert wird, der die größeren Ähnlichkeitsmaße liefert.
- 15 7. Gerät zum Zuteilen eines Benutzerkennzeichners, mit einem Spracherkenner und einer Steuereinrichtung, die zum Empfangen von Ausgangsdaten des Erkenners angeschlossen und betrieblich zur Durchführung des Verfahrens jedes der Ansprüche 1 bis 6 geschaltet sind.
8. Gerät nach Anspruch 7 zum Betreiben eines interaktiven, sprachgesteuerten Dienstes, der über eine Fernsprechleitung zugänglich ist.
- 20 9. Gerät nach Anspruch 8, bei dem der sprachbetriebene Dienst ein Home Banking-Dienst ist.

Revendications

- 25 1. Un procédé d'attribution à un utilisateur d'un service ou d'une prestation par commande vocale à un identificateur d'abonné comprenant une pluralité de mots provenant d'un jeu ou ensemble de mots, comprenant:
 - (i) la demande à l'utilisateur de prononcer les mots du jeu ou ensemble;
 - (ii) la comparaison de chaque mot prononcé avec un jeu de grilles d'identification vocale correspondant aux mots du jeu, pour obtenir une évaluation de leur similitude;
 - 30 (iii) la détermination, parmi les mots du jeu, de celui qui satisfait un critère sur l'étendue à laquelle l'évaluation de similitude du mot prononcé par rapport à la grille correspondant au mot demandé dépasse l'évaluation de similitude du mot prononcé par rapport à celui des autres grilles avec lequel la similitude du mot prononcé est la plus grande; et
 - (iv) la sélection d'une séquence de mots n'utilisant que des mots du jeu qui satisfont aux critères.
- 35 2. Un procédé selon la revendication 1, dans lequel le critère est que la différence de deux évaluations de similitude dépasse une valeur de seuil.
- 40 3. Un procédé selon la revendication 1 ou 2, dans lequel :
 - (a) chaque mot prononcé est comparé à deux jeux de grilles;
 - (b) les évaluations de similitude obtenues avec les deux jeux sont comparées pour identifier le jeu qui donne des évaluations qui sont, en général, plus grandes que celles de l'autre jeu; et
 - (c) la détermination concernant les mots qui satisfont le critère est effectuée en utilisant le jeu identifié.
- 45 4. Un procédé selon la revendication 1 ou 2, dans lequel:
 - (a) chaque mot prononcé est comparé à deux jeux de grilles;
 - (b) le jeu dans lequel les différences (D1-D2) entre les évaluations de similitude de chaque mot prononcé par rapport au mot demandé correspondant, et celui de l'une des autres grilles de chaque jeu avec le mot prononcé dont la similitude sont, en général, les plus grandes est identifié; et
 - 50 (c) la détermination des mots qui satisfont les critères est effectuée en utilisant le jeu identifié.
5. Un procédé selon la revendication 3 ou 4, dans lequel deux jeux de grilles correspondent, respectivement, à une voix masculine et féminine.
- 55 6. Un procédé selon la revendication 3, 4 ou 5, dans lequel, dans la sélection d'une séquence de mots, le premier des mots est pris à partir de l'un des deux sous-ensembles du jeu de mots, sur la base duquel on identifie celui des deux jeux de grilles comme donnant les valeurs les plus grandes.



7. Un appareil d'affectation d'un identificateur d'utilisateur comprenant un identificateur vocal et des moyens de commande connectés pour recevoir les données de sortie de l'identificateur et dont le fonctionnement est agencé pour exécuter le procédé de l'une quelconque des revendications 1 à 6.
8. Un appareil selon la revendication 7 pour effectuer un service interactif commandé par la voix, accessible par l'intermédiaire d'une ligne téléphonique.
9. Un appareil selon la revendication 8, dans lequel le service commandé par la voix est un service bancaire domestique.

10

15

20

25

30

35

40

45

50

55

