



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁵ : G09B 19/06</p>	<p>A1</p>	<p>(11) International Publication Number: WO 90/01203 (43) International Publication Date: 8 February 1990 (08.02.90)</p>
<p>(21) International Application Number: PCT/GB89/00846 (22) International Filing Date: 25 July 1989 (25.07.89) (30) Priority data: 8817705.0 25 July 1988 (25.07.88) GB (71) Applicant (for all designated States except US): BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB). (72) Inventor; and (75) Inventor/Applicant (for US only) : STENTIFORD, Frederick, Warwick, Michael [GB/GB]; Sheepstor, Boyton, Woodbridge, Suffolk IP12 3LH (GB). (74) Agent: LLOYD, Barry, George, William; British Telecommunications public limited company, Intellectual Property Unit, 151 Gower Street, London WC1E 6BA (GB).</p>		<p>(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), DK, FI, FR (European patent), GB, GB (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), NO, SE (European patent), SU, US.</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>

(54) Title: LANGUAGE TRAINING

(57) Abstract

A speech synthesizer (3) produces prompts in the voice of a native speaker of the language to be learned, which the student may imitate or reply to, and a phrase recogniser (1) which uses keyword recognition is employed so that the system understands spoken phrases and interactive dialogue may take place. The student's progress is monitored by measuring the deviation from his original speech recognition template; when this difference is sufficiently large that the recogniser (1) can no longer recognise what the student is saying, the system re-trains and updates the template. In another embodiment, the system includes a display which shows the native speaker's mouth shape whilst the words to be imitated are spoken by the speech synthesizer (3); and a video pick-up and analyser for analysing the shapes of the student's mouth to give the student visual feedback.

```

graph TD
    A[INPUT SOURCE & TARGET LANGUAGE ?] --> B[INPUT SUBJECT AREA ?]
    B --> C[OUTPUT NATIVE SPEAKER KEYWORD PROMPTS & RECORD TRAINEE INPUTS]
    C --> D[USE TRAINEE INPUTS TO FORM TEMPLATES]
    D --> E[INDICATE IMPROVEMENT]
    D --> F[INDICATE IMPROVEMENT]
    E --> G[OUTPUT SOURCE LANGUAGE WORD OR PHRASE FOR TRANSLATION]
    G --> H[INPUT TARGET LANGUAGE RESPONSE]
    H --> I{DOES INPUT DIFFER FROM TEMPLATE MUCH ?}
    I -- YES --> D
    I -- NO --> J[ ]
    F --> K[OUTPUT TARGET LANGUAGE DIALOGUE]
    K --> L[INPUT TARGET LANGUAGE RESPONSE]
    L --> M{DOES INPUT DIFFER FROM TEMPLATE MUCH ?}
    M -- YES --> D
    M -- NO --> N[ ]
    style J fill:none,stroke:none
    style N fill:none,stroke:none
    
```

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Faso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	HU	Hungary	NO	Norway
BJ	Benin	IT	Italy	RO	Romania
BR	Brazil	JP	Japan	SD	Sudan
CA	Canada	KP	Democratic People's Republic of Korea	SE	Sweden
CF	Central African Republic	KR	Republic of Korea	SN	Senegal
CG	Congo	LI	Liechtenstein	SU	Soviet Union
CH	Switzerland	LK	Sri Lanka	TD	Chad
CM	Cameroon	LJ	Luxembourg	TG	Togo
DE	Germany, Federal Republic of	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

LANGUAGE TRAINING

This invention relates to apparatus and methods for training pronunciation; particularly, but not exclusively, for training the pronunciation of second or foreign languages.

5 One type of system used to automatically translate speech between different foreign languages is described in our European published patent application number 0262938A. This equipment employs speech recognition to
10 recognise words in the speaker's utterance, pattern matching techniques to extract meaning from the utterance and speech coding to produce speech in the foreign tongue.

This invention uses similar technology, but is configured in a different way and for a new purpose, that
15 of training a user to speak a foreign language.

This invention uses speech recognition not only to recognise the words being spoken but also to test the consistency of the pronunciation. It is a disposition of
20 novice students of language that, although they are able to imitate a pronunciation, they are liable to forget, and will remain uncorrected until they are checked by an expert. A machine which was able to detect mispronunciation as well as translation inaccuracies would
25 enable students to reach a relatively high degree of proficiency before requiring the assistance of a conventional language teacher to progress further. Indeed, very high levels of linguistic skill are probably
30 not required in the vast majority of communication tasks, such as making short trips abroad or using the telephone, and computer aided language training by itself may be sufficient in these cases.

Conventional methods either involve expensive skilled human teachers, or the use of passive recordings of foreign speech which do not test the quality of the student's pronunciation.

5 Some automated systems provide a visual display of a representation of the student's speech, and the student is expected to modify his pronunciation until this display matches a standard. This technique suffers from the disadvantage that users must spend a great deal of time
10 experimenting and understanding how their speech relates to the visual representation.

 Another approach (described for example in *Revue de Physique Appliquee* vol 18 no. 9 Sept 1983 pp 595-610, M.T. Janot-Giorgetti et al, "Utilisation d'un systeme de
15 reconnaissance de la parole comme aide a l'acquisition orale d'une langue etrangere") employs speaker independent recognition to match spoken utterances against standard templates. A score is reported to the student indicating how well his pronunciation matches the ideal. However,
20 until speaker independent recognition technology is perfected, certain features of the speaker's voice, such as pitch, can affect the matching scores, and yet have no relevant connection with the quality of pronunciation. A student may therefore be encouraged to raise the pitch of
25 his voice to improve his score, and yet fail to correct an important mispronunciation.

 Furthermore, current speaker independent recognition technology is unable to handle more than a small vocabulary of words without producing a very high error
30 rate. This means that training systems based on this technology are unable to process and interpret longer phrases and sentences. A method of training pronunciation for deaf speakers is described in *Proceedings ICASSP 87* vol 1 pp 372-375 D. Kewley-Port et al 'Speaker-dependant

Recognition as the Basis for a Speech Training Aid'. In this method, a clinician selects the best pronounced utterances of a speaker and these are converted into templates. The accuracy of the speaker's subsequent pronunciation is indicated as a function of his closeness to the templates (the closer the better). This system has two disadvantages; firstly, it relies upon human intervention by the clinician, and secondly the speaker cannot improve his pronunciation over his previous best utterances but only attempt to equal it.

According to the invention there is provided apparatus for pronunciation training comprising;

- speech generation means for generating utterances; and

- speech recognition means arranged to recognise in a trainee's utterances, the words from a predetermined selected set of words,

wherein the speech recognition means is arranged to employ speaker-dependent recognition, by comparing the trainee's utterance with templates for each word of the set, and the apparatus is arranged initially to generate the templates by prompting the trainee to utter each word of the set and forming the templates from such utterances, the apparatus being further arranged to indicate improvements in pronunciation with increases in the deviation of the trainee's subsequent utterances from the templates.

Some non-limitative examples of embodiments of the invention will now be described with reference to the drawings, in which:

- Figure 1 illustrates stages in a method of language training according to one aspect of the invention;

- Figure 2 illustrates schematically apparatus suitable for performing one aspect of the invention;

- Figure 3 illustrates a display in an apparatus for language training according to another aspect of the invention.

5 Referring to Figures 1 and 2, upon first using the system illustrated, the student is asked by the system (using either a screen and keyboard or conventional speech synthesiser and speaker independent recogniser) which language he wishes to study, and which subject area (eg operating the telephone or booking hotels) he requires.
10 The student then has to carry out a training procedure so that the speaker dependent speech recogniser 1 can recognise his voice. To this end, the student is prompted in the foreign language by a speech generator 3 employing a pre-recorded native speaker's voice to recite a set of
15 keywords relevant to his subject area. At the same time, the source language translation of each word is displayed, giving the student the opportunity to learn the vocabulary. This process, in effect, serves as a passive learning stage during which the student can practise his
20 pronunciation, and can repeat words as often as he likes until he is satisfied that he has imitated the prompt as accurately as he believes he can.

A control unit 2 controls the sequence of prompts and responses. Conveniently, the control unit may be a
25 personal computer (for example, the IBM PC).

These utterances are now used as, or to generate, the first set of templates stored in template store 1a to be used by the speech recogniser 1 to process the student's voice. The templates represent the student's first
30 attempt to imitate the perfect pronunciation of the recorded native speaker.

The second stage of the training process simply tests the ability of the student to remember the translations and pronunciations of the key word vocabulary. He is

prompted in his source language (either visually, on screen 4, or verbally by speech generator 3) to pronounce translations of the keywords he has practised in the previous stage. After each word is uttered, the speech generator 3 repeats the foreign word recognised by the recogniser 1 back to the student and displays the source language equivalent. Incorrect translations are noted for re-prompting later in the training cycle. The student is able to repeat words as often as he wishes, either to refine his pronunciation or to correct a machine misrecognition. If the recogniser 1 consistently (more than, say, 5 times) misrecognises a foreign word, either because of a low distance score or because two words are recognised with approximately equal distances, the student will be asked to recite this word again (preferably several times), following a native speaker prompt from the generator 3, so that a new speech recognizer template can be produced to replace the original template in store 1a. Such action in fact indicates that the student has changed his pronunciation after having heard the prompt several more times, and is converging on a more accurate imitation of the native speaker. This method has the advantage over the prior art that the trainee's progress is measured by his deviation from his original (and/or updated) template, rather than by his convergence on the native speaker's template, thus eliminating problems due to pitch, or other, differences between the two voices. Once the student is satisfied that he has mastered the key word vocabulary, he may move to the third training stage.

The student is now prompted in his own language (either visually on screen 4 or verbally through generator 3) and may be asked to carry out verbal translations of words or complete phrases relevant to his subject area of interest. Alternatively, these prompts may take the form

of a dialogue in the foreign language to which the student must respond. One useful method of prompting is a 'storyboard' exercise using a screen display of a piece of text, with several words missing, which the student is prompted to complete by uttering what he believes are the missing words. The system now preferably operates in the same manner as the phrase-based language translation system (European Published Application No 0262938) and recognises the pre-trained keywords in order to identify the phrase being uttered. The system then enunciates the correct response/translation back to the student in a native speaker's voice, and gives the student an opportunity to repeat his translation if it was incorrect, if he was not happy with the pronunciation, or if the recogniser 1 was unable to identify the correct foreign phrase. In the event that the student is unable to decide whether the recogniser 1 has assimilated his intended meaning, the source language version of the recognised foreign phrase can be displayed at the same time. Incorrectly translated phrases are re-presented (visually or verbally) to the student later in the training cycle for a further translation attempt.

If the recogniser 1 repeatedly fails to identify the correct phrase because of poor key word recognition and drifting student pronunciation, the student will be asked to recite each key word present in the correct translation for separate recognition. If one or more of these keywords is consistently misrecognised, new templates are generated as discussed above.

Phrases are presented to the student for translation in an order which is related to their frequency of use in the domain of interest. The system preferably enables the trainee to suspend training at any point and resume at a

later time, so that he is able to progress as rapidly or as slowly as he wishes.

5 The preferred type of phrase recognition (described in European Published Application No 0262938 and 'Machine Translation of Speech' Stentiford & Steer, British Telecom Technology Journal Vol 6 No. 2 April '88 pp 116-123) requires that phrases with variable parameters in them such as dates, times, places or other sub-phrases, should be treated in a hierarchical manner. The form of the
10 phrase is first identified using a general set of keywords. Once this is done, the type of parameter present in the phrase can be deduced and a special set of keywords applied to identify the parameter contents. Parameters could be nested within other parameters. As a
15 simple example, a parameter might refer to a major city in which case the special keywords would consist of just these cities. During student training translation, errors in parameter contents can also be treated hierarchically. If the system has identified the correct form of phrase
20 spoken by the student, but has produced an incorrect parameter translation, the student can then be coached to produce the correct translation of the parameter in isolation, without having to return to the complete phrase.

Parameters are normally selected in a domain of
25 discourse because of their occurrence across a wide range of phrases. It is natural therefore that the student should receive specific training on these items if he appears to have problems with them.

The keywords are selected according to the information
30 they bear, and how well they distinguish the phrases used in each subject area. This means that it is not necessary for the system to recognise every word in order to identify the phrase being spoken. This has the advantage that a number of speech recognition errors can be

tolerated before phrase identification is lost. Furthermore, correct phrases can be identified in spite of errors in the wording which might be produced by a novice. It is reasonable to conjecture that, if the
5 system is able to match attempted translations with their corrected versions, such utterances should be intelligible in practice when dealing with native speakers who are aware of the context. This means that the system tends to concentrate training on just those parts of the student's
10 diction which give rise to the greatest ambiguity in the foreign language. This might be due to bad pronunciation of important keywords or simply due to their omission.

The described system therefore provides an automated learning scheme which can rapidly bring language students
15 up to a minimum level of intelligibility, and is especially useful for busy businessmen who simply wish to expedite their transactions, or holiday makers who are not too worried about grammatical accuracy.

The correct pronunciation of phrases is given by the
20 recorded voice of a native speaker, who provides the appropriate intonation and co-articulation between words. The advanced student is encouraged to speak in the same manner, and the system will continue to check each utterance, providing the word spotting technology employed
25 is able to cope with the increasingly fluent speech.

Referring to Figure 3, in another aspect of the invention, a visual display of the mouth of the native speaker is provided so as to exhibit the articulation of each spoken phrase. This display may conveniently be
30 provided on a CRT display using a set of quantised mouth shapes as disclosed in our previous European Published Application No. 0225729A. A whole facial display may also be used.

In one simple embodiment, the display may be mounted in conjunction with a mirror so that the applicant may imitate the native speaker.

5 In a second embodiment, a videophone coding apparatus of the type disclosed in our previous European Published Application No. 0225729 may be employed to generate a corresponding display of the student's mouth so that he can accurately compare his articulation with that of the native speaker. The two displays may be simultaneously
10 replayed by the student, either side by side, or superimposed (in which case different colours may be employed), using a time-warp method to align the displays.

15

CLAIMS

1. Apparatus for pronunciation training comprising;
 - speech generation means for generating utterances; and
 - 5 - speech recognition means arranged to recognise in a trainee's utterances, the words from a predetermined selected set of words,
 - wherein the speech recognition means is arranged to employ speaker-dependent recognition, by comparing the
 - 10 trainee's utterance with templates for each word of the set, and the apparatus is arranged initially to generate the templates by prompting the trainee to utter each word of the set and forming the templates from such utterances, the apparatus being further arranged to indicate
 - 15 improvements in pronunciation with increases in the deviation of the trainee's subsequent utterances from the templates.

2. Apparatus according to claim 4, further arranged to update the templates from the said subsequent
- 20 utterances when the said deviation exceeds a predetermined threshold.

3. Apparatus according to claim 1 or claim 2 further comprising control means connected to the speech generation means and to the speech recognition means, and
- 25 arranged so anticipated that, in use, the apparatus generates a prompt to which a trainee may respond by speaking, the speech recognition means is arranged to recognise in the trainee's response the presence of words from the said set, and the speech generation means is
- 30 arranged to generate an utterance in dependence on what the speech recognition means has recognised.

- 11 -

4. Apparatus for pronunciation training according to claim 3, further comprising;

5 - phrase recognition means for identifying phrases by the combination and order of words from the said predetermined selected set,

10 wherein in use the trainee is prompted to respond by uttering a phrase, the phrase recognition means recognises the phrase and the utterance generated by the speech generation means is thereby selected to be a reply to the phrase.

5. Apparatus for pronunciation training according to claim 3 or claim 4, wherein the prompt is an utterance generated by the speech generation means.

15 6. Pronunciation training apparatus comprising speech generation means for generating utterances, and video generation means for generating corresponding video images of a mouth, whereby a trainee is prompted to imitate the correct pronunciation of the said utterances.

20 7. Apparatus according to claim 6, further comprising video analysis means arranged to analyse mouth movements of the trainee and to display the corresponding synthesised and analysed mouth movements.

25 8. Language training apparatus according to any preceding claim, wherein the speech generation means is arranged to generate utterances in a language in the accent of a native speaker of that language.

9. A method of pronunciation training, comprising;
- prompting a trainee to speak an utterance, and

- analysing the utterance using speaker-dependent speech recognition, employing templates derived from the trainee's previous utterances;

5 whereby improvements in pronunciation are assessed by measuring the distance between the utterance and the template; the assessment being such that an increase in distance corresponds to a pronunciation improvement.

10 10. A method according to claim 9 further comprising the step of: updating the said templates when the said distance exceeds a predetermined threshold.

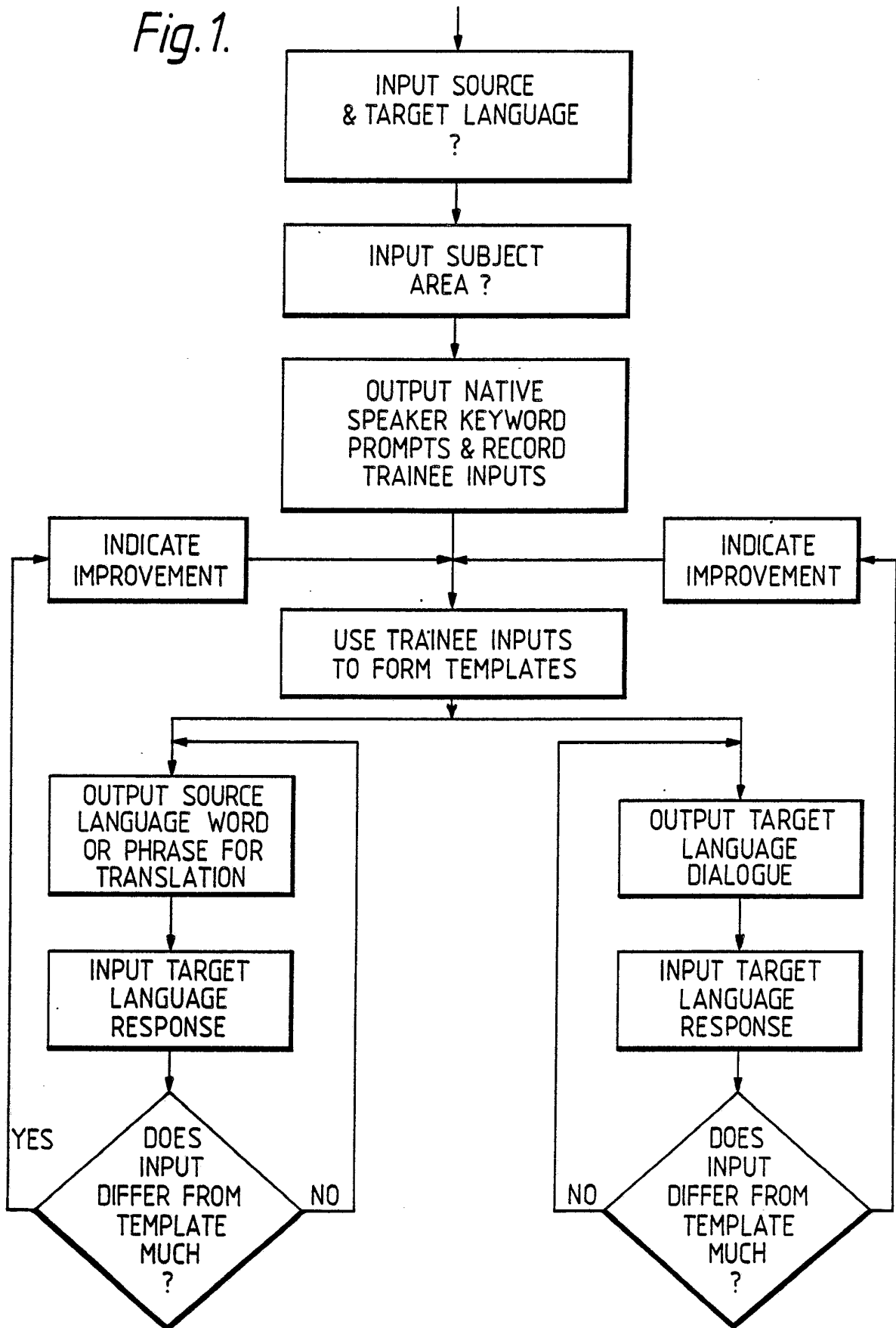
11. A method of pronunciation training comprising employing apparatus according to any one of claims 1 to 6.

15 12. Apparatus for pronunciation training substantially as herein described with reference to Figure 1 and Figure 2, or Figure 3.

13. A method of pronunciation training substantially as herein described with reference to Figure 1 or Figure 3.

7/2

Fig. 1.



2/2

Fig. 2.

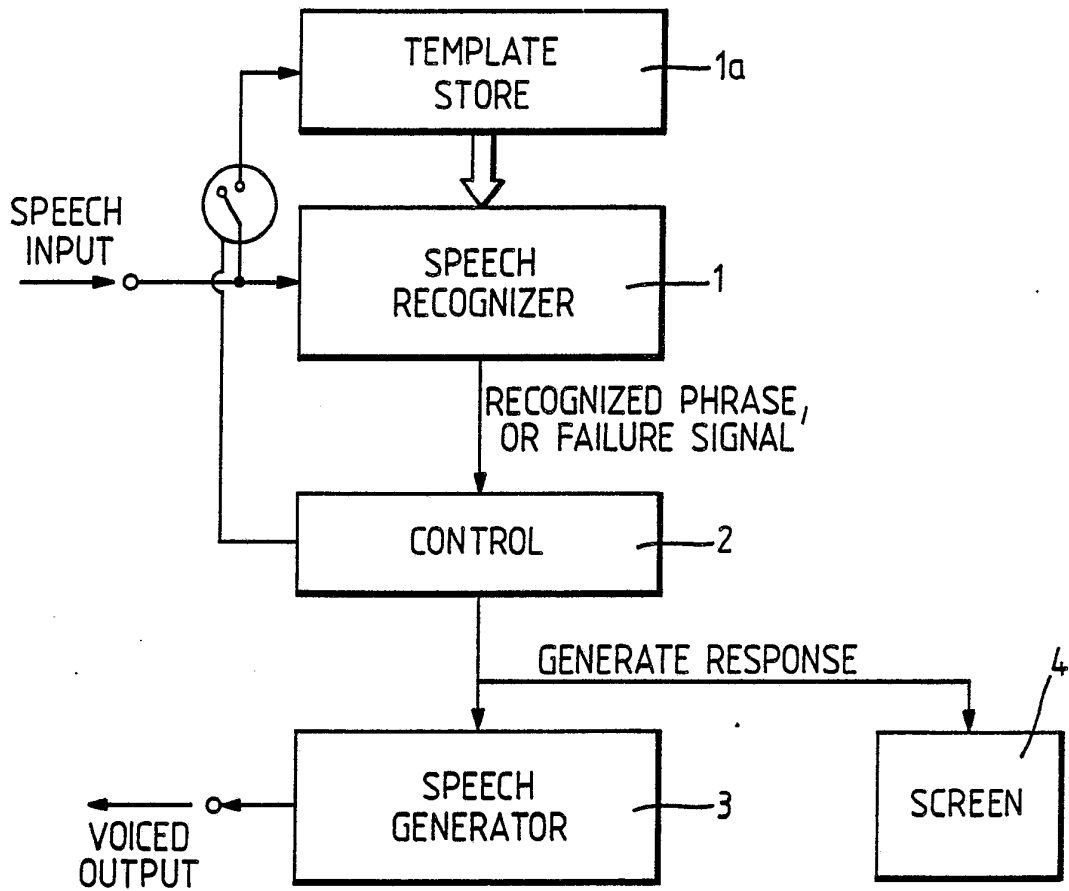
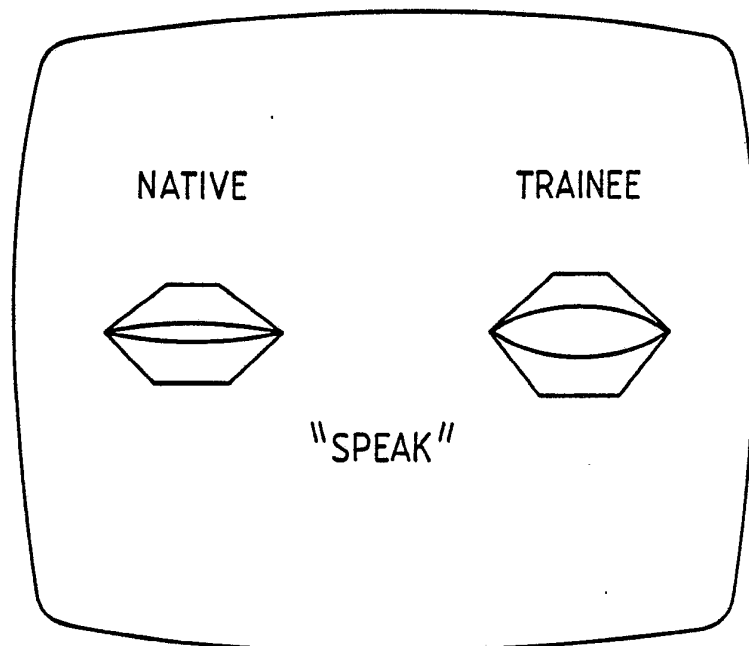
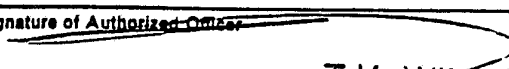


Fig. 3.



INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 89/00846

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ⁶		
According to International Patent Classification (IPC) or to both National Classification and IPC		
IPC ⁵ : G 09 B 19/06		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
IPC ⁵	G 09 B	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁸		
III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹		
Category ⁹	Citation of Document, ¹¹ with Indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
Y	Revue de Physique Appliquée, vol. 18, no. 9, September 1983, Orsay (FR) M.T. Janot-Giorgetti et al.: "Utilisation d'un système de reconnaissance de la parole comme aide à l'acquisition orale d'une langue étrangère" pages 595-610, see pages 597, 598, 606-608	1, 3, 4, 9
A	(cited in the application)	8, 11, 12, 13
Y	EP, A, 0094502 (BARBARA THOMPSON) 23 November 1983, see claims	1, 3, 4, 9
A	--	5, 8, 11, 13
Y	EP, A, 0262938 (FREDERICK WARWICK MICHAEL STENTIFORD) 6 April 1988, see columns 1, 2; figure	1, 3, 4, 9
A	(cited in the application)	8, 11, 12
	--	
<p>¹⁰ Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"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</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"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.</p> <p>"&" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search 11th December 1989	Date of Mailing of this International Search Report 19. 01. 90	
International Searching Authority EUROPEAN PATENT OFFICE	Signature of Authorized Officer  T.K. WILLIS	

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category*	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	<p>Proceedings ICASSP 87 International Conference on Acoustics, Speech, and Signal Processing, Dallas, 6-9 April 1987, vol. 1, IEEE (US)</p> <p>D. Kewley-Port et al.: "Speaker-dependent speech recognition as the basis for a speech training aid", pages 372-375 see pages 372,373</p> <p>(cited in the application)</p> <p style="text-align: center;">--</p>	1,3,4,9
A	<p>EP, A, 0225729 (WILLIAM JOHN WELSH et al.)</p> <p>16 June 1987, see pages 23-27; claims 1-8,15,16,20; figures 1-3</p> <p>(cited in the application)</p> <p style="text-align: center;">--</p>	6,7
A	<p>GB, A, 2167224 (ROBERT SPRAGUE et al.)</p> <p>21 May 1986, see claims; figures 1-9</p> <p style="text-align: center;">-----</p>	6,7

**ANNEX TO THE INTERNATIONAL SEARCH REPORT
ON INTERNATIONAL PATENT APPLICATION NO.**

GB 8900846
SA 30463

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 09/01/90. The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0094502	23-11-83	JP-A- 59026799	13-02-84
EP-A- 0262938	06-04-88	WO-A- 8802516	07-04-88
EP-A- 0225729	16-06-87	JP-A- 62120179 US-A- 4841575	01-06-87 20-06-89
GB-A- 2167224	21-05-86	US-A- 4650423 US-A- 4795349 US-A- 4768959	17-03-87 03-01-89 06-09-88