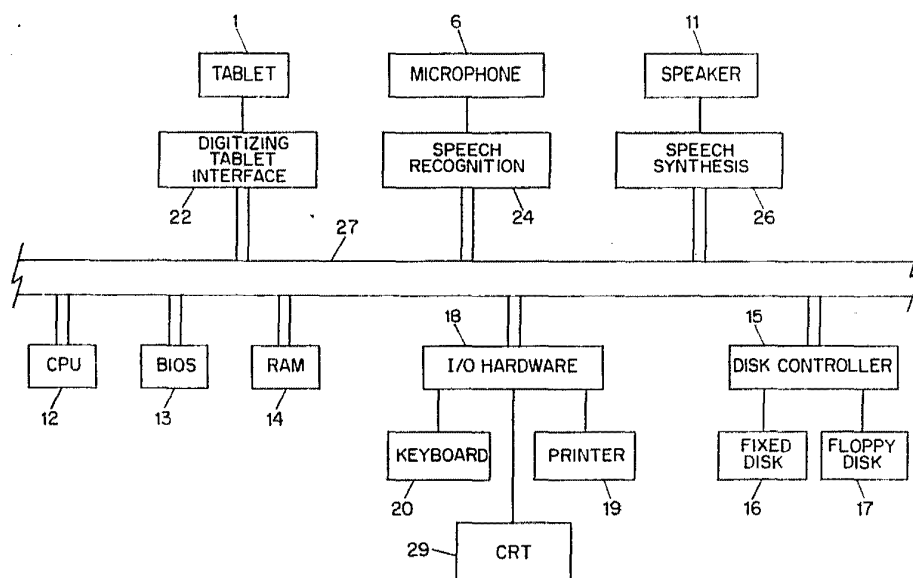




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : G06K 9/22, G06F 3/16	A1	(11) International Publication Number: WO 92/05517 (43) International Publication Date: 2 April 1992 (02.04.92)
(21) International Application Number: PCT/US91/06874 (22) International Filing Date: 23 September 1991 (23.09.91) (30) Priority data: 588,185 26 September 1990 (26.09.90) US (71)(72) Applicant and Inventor: ROTH, Richard, G. [US/US]; 8 Sutton Court, Great Neck, NY 11021-1318 (US). (74) Agents: MAUNE, James, J. et al.; Brumbaugh, Graves, Donohue & Raymond, 30 Rockefeller Plaza, Suite 44, New York, NY 10112 (US). (81) Designated States: AT (European patent), AU, BE (European patent), BG, BR, CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), GR (European patent), HU, IT (European patent), JP, KR, LU (European patent), NL (European patent), NO, RO, SE (European patent), SU ⁺ .		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>

(54) Title: AUDIO-AUGMENTED HANDWRITING RECOGNITION**(57) Abstract**

An improved data-entry apparatus for entering handwritten data is disclosed. The user enters a word (or portion thereof, depending on the language and notation used) into a touchpad and speaks the word. Features are extracted from the spoken information. The touchpad data are indicative of any of a number of candidate words, and a multiplicity of speech portion templates are generated, each template indicative of one of the candidate words. The apparatus evaluates the correlation between the extracted features and the features of each generated speech portion template, and the speech portion template having the highest correlation with the extracted features determines the recognized word. Optionally a speech synthesizer speaks the word and an opportunity is provided so the user may confirm the correctness of the recognition.

⁺ See back of page

+ DESIGNATIONS OF "SU"

Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

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	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU ⁺	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

Description

Audio-Augmented Handwriting Recognition

Background of the Invention

This invention relates generally to apparatus for receiving handwritten (calligraphic) input, and relates more particularly to a novel technique for augmenting the handwritten input with audio (spoken) input.

5 Much attention has been paid in recent decades to the prospect of computerized recognition of handwritten input. Typically an input device such as a touchpad is employed, which receives user input by a stylus (or
10 finger) and provides information about the position of the stylus in an orthogonal reference frame. Provided to the recognition system is information indicative of the position, generally in X-Y coordinates, of the stylus and whether or not the stylus is touching the touchpad. With other touchpad designs the recognition
15 system receives no signal unless the stylus is touching the touchpad; when the stylus is touching, signals provide the X and Y coordinates.

In some systems the information is stored for later analysis, in which case the task becomes closely
20 analogous to ordinary optical character recognition. In other systems the information from the touchpad is analyzed at or near real time, and the information analyzed may include not only the X and Y coordinates of the stylus but also temporal information regarding
25 the number of strokes making up the character, the order of the strokes, the direction of the writing for each stroke, and even the speed of the writing within each stroke. In a few systems the stylus is able to resolve several degrees of contact pressure.

30 Few if any of the known on-line character recognition systems are really quite satisfactory. A

few systems are fairly successful but require computing power far in excess of that available in hardware on the scale of a personal computer (PC). To achieve high accuracy of recognition it is generally necessary to
5 constrain closely the permitted range of inputs, such as to the numerical digits 0 and 1 through 9. Where the range of possible inputs is intended to include, say, a Roman alphabet, most systems have poor recognition accuracy or constrain closely the type of
10 letters (e.g. capital, block letters) that are entered.

Other systems can only achieve acceptable accuracy after having the opportunity to "learn" the handwriting of a particular user, and do not perform well with arbitrary users whose handwriting has not been
15 "learned".

One particularly large factor in the limited success of most handwriting recognition systems is that the system either recognizes a character or it does not. That is, the algorithm that accepts the touchpad
20 input information and yields a recognized character faces the design constrain of coming up with a single character result, or none at all. And the result has to be correct most of the time or the overall system will be of little utility. The requirement that the
25 single result must be right most of the time dictates, in most systems, one or another of the compromises mentioned above: giving up on reading everyone's handwriting, giving up on reading a wide range of characters, giving up on reading both upper- and lower-
30 case letters, or giving up on obtaining the results in real time.

Summary of the Invention

In accordance with the invention there is provided an improved data-entry apparatus for entering
35 handwritten data. The user enters a speech portion (a word, a part of a word, or multi-word phrase) by

writing it on a tablet and speaking it into a microphone. Features are extracted from the spoken information. The touchpad data are indicative of any of a number of candidate words, and a multiplicity of
5 speech portion templates are generated, each template indicative of one of the candidate words. The apparatus evaluates the correlation between the extracted features and the features of each generated speech portion template, and the speech portion
10 template having the highest correlation with the extracted features determines the recognized word. Optionally a speech synthesizer speaks the word, or the recognition result is provided to the user on a visual display, and an opportunity is provided so the user may
15 confirm the correctness of the recognition.

Among the salient features of the apparatus of the invention is that the recognition algorithm need not satisfy the standard of yielding a single recognition result to the exclusion of all other results. Instead,
20 the algorithm is permitted to return (upon analysis of the touchpad input) with a list of candidates. Pronunciation-by-rule algorithms are employed to arrive at a pronunciation (or in the case of allophones, a plurality of pronunciations) for each candidate.
25 Speech recognition features are extracted from the voice input of the user, who will have spoken the word as well as having entered it at the touchpad. The features are compared with the pronunciations, and the nearest match is used as the result of the recognition.
30 Optionally, the system may synthesize and replay for the user the result, giving the user a chance to indicate whether or not the recognition was unsuccessful, either by spoken response or by touchpad entry.

Brief Description of the Figures

The invention will be described in more detail by a drawing, of which:

Fig. 1 is a diagram, in functional-block form, of the Audi- augmented handwriting recognition device of the invention;

Fig. 2 is a system configuration diagram, showing the system according to the invention.

Detailed Description

Fig. 1 shows in functional block diagram form the system according to an embodiment of the invention.

Handwritten input is provided by the user at touchpad 1, which preferably has resolution in each of two axes of at least 200 divisions per inch. Data regarding stylus position is collected as indicated by block 2 as a function of time, preferably at least as often as 100 samples per second. Where the invention is applied to recognition of, say, individual letters in a Roman alphabet, the user entry is typically less than a second, and only a few kilobytes (K) of data are collected. As will be discussed further below, the invention shows promise for recognizing Chinese characters, in connection with which the data storage may be tens or hundreds of K bytes. Also discussed below is the case where the invention is applied to alphabetic entry of entire words or phrases, in which case the data storage may be at least several tens of K bytes.

Once the data are collected, the data points are configured as indicative of the strokes making up the character to be recognized. Several known categories of preprocessing are performed, symbolized by block 3. These preferably include smoothing of strokes, filtering of stroke data, correcting wild points such as outlying points in the data, dehooking (removing idiosyncratic movements at ends of strokes), reducing

multi-point dots of stylus movement to a single dot,
and stroke correction in cases where a spurious
indication of, say, lifting the stylus from the pad
wrongly results in a single stroke being recorded as
5 two strokes.

Prior to the attempt to recognize the character a
number of types of normalization are preferably also
performed as part of the pre- recognition processing.
These include the known types of normalization such as
10 deskewing, baseline drift correction, size
normalization, stroke length normalization, all known
to those skilled in the art.

In prior art handwriting recognition systems the
next step, recognition, is quite difficult. Known
15 recognition methodologies which have been employed
include feature matching, in which features of the to-
be-recognized character are tested for match with
characters in a feature database; time sequence of
zones, directions, or extremes, where a simple database
20 match, or a binary tree analysis, may be employed
regarding relatively discrete information about the
character-formation temporal data; curve matching, in
which the curves making up the character are compared
with standard curves in a database; or stroke codes,
25 where parts of a character are classified and
enumerated, then compared with a database. As
mentioned above, all these methodologies suffer, when
used in prior art handwriting recognition systems, from
the problem that they are of no use unless they give a
30 single highly unambiguous result for each recognized
character. To achieve this singular result, it is
generally required that the size of the character set
be constrained, that the font be constrained, or that
the number of users be constrained.

35 At block 4 of Fig. 1, the recognition is
performed, but the recognition algorithm only requires

that the range of possible recognition results be confined to a small number of matches. Where letters of an alphabet are being recognized, the number of candidates can be half a dozen or so. Where Chinese
5 characters are being recognized, the number of candidates may typically be a few dozens or a few hundred.

The next stage of processing, shown as block 5, is the employment of pronunciation-by-rule methods to
10 arrive at a pronunciation for each of the candidates. Some letters or characters may have more than one pronunciation, and if so both pronunciations are arrived at. Such letters (or other speech portions) are called allophones, and do not cause a problem for
15 the apparatus of the invention. Feature templates for each of the pronunciations are stored; the number of feature templates to be compared with the spoken features may thus exceed the number of candidate speech portions. Similarly, it may be desired to permit, say,
20 American and British templates for speech portions pronounced differently in the two dialects, or for speech portions pronounced by men or women.

At about the same time, the user is asked to speak the character into a microphone. Well-known speech
25 recognition steps are taken, starting with storage of the spoken information, preferably in RAM. A few megabytes suffice for storage of several seconds of speech, collected by microphone 6 and audio storage means 7. Features are extracted by known means, such
30 as determination of the amount of speech energy within various frequency ranges as a function of time.

In accordance with the invention, the novel step of correlation is performed. The extracted features of the spoken information are compared one by one with the
35 features of the candidates mentioned before. This step, shown as functional block 9, assumes a meaningful

correlation measure which, in the general case, can be expected to permit selection of one of the candidates (call it candidate #1) as being a closer feature match than the others.

5 In order for this correlation to result in an "acceptance", two requirements must be met:

 (1) the extracted features of the spoken information must match the features of candidate #1 with a correlation value exceeding some predetermined
10 threshold (i.e. it must be a "sufficiently close" match); and

 (2) the extracted features of the spoken information must match the features of candidate #1 with a correlation value exceeding the correlation
15 values for the other candidates by some predetermined margin (i.e. it must be a "better" match than all the runners-up).

 Experience suggests that with properly chosen threshold and margin (and, of course, with a workable
20 correlation measure) then most of the time the two requirements will be met, and recognition will have been accomplished on the "first pass". However, according to the invention any of several remedial measures may be taken during a so-called "second pass".
25 For example, the system may:

 (1) offer a number of candidates, ranked by correlation value, for multiple-choice selection by the user; or

 (2) have the user repeat the reciting and writing
30 process; or

 (3) where an alphabet has not yet been used but is available, have the user enter the information character by character.

 In the case of Kanji entry, one example of the
35 last option is having the user enter the information by kana. In the case of entry of a cursive English word,

one example of the last option is having the user enter the word letter by letter.

Optionally, each feature match is followed as in block 10 with an audio synthesis of the match character. The synthesized speech is replayed for the user, and assuming the user agrees with the match, the process continues with the next handwriting to be recognized. In the (hopefully rare) case that the match was incorrect, the user will so indicate, either by touchpad input or by speech input to the system.

Instead of or in addition to an audio synthesis of the match character, there may be a visual display of the match character as indicated by block 28. Assuming the user agrees with the match, the process continues with the next handwriting to be recognized. As with the audio synthesis technique just described, in the (hopefully rare) case that the match was incorrect, the user will so indicate, either by touchpad input or by speech input to the system.

Turning now to Fig.2 there is shown a block diagram of a system according to the invention based on a known personal computer. CPU 12, preferably a fast processor such as a high-speed 80386 microprocessor, forms an AT-compatible or EISA-compatible system with BIOS 13, fixed disk 16 (and disk controller 15), optional floppy disk 17, RAM 14, and I/O interface hardware 18 to drive keyboard 20, display 29, and printer 19. All the above-mentioned are in communication by means of bus 27.

Provided as peripherals to the microcomputer are touchpad tablet 1 with interface 22 plugged into an expansion slot of bus 27, microphone 6 with speech recognition (feature extraction) interface 24, and an optional speaker with speech synthesis card 26. All the three expansion peripherals comprise known hardware, available off-the-shelf.

The steps depicted functionally in Fig. 1, blocks 2 through 5 and 7 through 9 (and optionally 10 and 28) are all performed essentially by the CPU 12 under program control. That is, nearly all the steps are
5 performed in software.

With data entry rates that are typical for humans (several tenths of a second for Roman characters, a second or two for Chinese characters) theory suggests the bandwidth of a fast AT or EISA bus and
10 microprocessor throughput can permit recognition according to the invention not only for individual letters but also for Chinese characters or entire words in Roman-alphabet languages.

A variant of the preferred embodiment is possible
15 in which, instead of pronunciation-by-rule, a template store is employed. The candidates obtained by handwriting recognition are pointers into a template store (preferably in fixed disk 16) containing speech templates, one (or more, in the case of allophones) for
20 each candidate. Once the template entries corresponding to the candidates have been identified, the correlation step is performed, to find out which entry (and thus which handwriting candidate) is the closest match to the spoken input.

25 Note that the embodiment just described has some elements in common with known speech recognition systems. But the embodiment offers many advantages over such known systems. For example, it will be appreciated that some prior art speech recognition
30 systems try to match the received features with every entry in a template store. In the case where the template store is large, with perhaps many tens of thousands of entries, this requires many tens of thousands of comparisons and is fraught with the danger
35 of an incorrect result since several entries may turn out to have correlation coefficients not far below that

of the entry with the highest correlation coefficient.
In the case where the template store is small, on the
other hand, the prospect of having one entry correlate
much better than the others is greater but the obvious
5 drawback is the limited recognition vocabulary.

Thus one advantage of the inventive embodiment may
be seen -- where analysis begins with the handwritten
input the portion of the template store that must be
compared with the received features is reduced
10 substantially. This saves computational time and
enhances the prospect of correct recognition on the
first try.

A hybrid embodiment is also possible, in which the
candidate words (or other speech portions) are used as
15 pointers into the template store for those words which
happen to be in the template store; for other words
(namely for words not having entries in the template
store) then pronunciation-by-rule is used to derive
pronunciation templates. The features extracted from
20 the spoken input are compared to template candidates
from either or both of the sources, namely from the
template store or from the pronunciation-by-rule
activity. As with the previous systems, a "best match"
is found which preferably matches more closely by a
25 predetermined margin than any of the other matches.

It will be appreciated that it is not strictly
necessary first to store the correlation coefficients
for all of the candidates, and only later to determine
which candidate had the highest coefficient (and thus
30 the best match). Instead, one approach is to establish
and maintain only two coefficients, each initialized to
zero. One is the "best" coefficient found thus far in
the comparison process, and the other is the "second
best" coefficient found thus far. Only if a particular
35 comparison yields a coefficient better than either the

-11-

"best" or "second Best" found thus far will it be necessary to update one or the other, or both.

It will be noted that the above-described embodiment is disclosed with respect to one particular sequence, namely (1) analyzing the handwritten input to
5 derive a number of candidate writing portions (e.g. words) and (2) receiving spoken input to disambiguate (e.g. narrow down the candidates to one result). This is the preferred embodiment because experience suggests
10 the present state of technology favors recognition of a relatively high quality graphical input; speech recognition, over the anticipated wide vocabulary to which the invention would be applied, is less reliable in and of itself, though it is thought sufficiently
15 workable for the above-mentioned disambiguation.

The invention may be optionally embodied with the opposite sequence, namely (1) analyzing the spoken input to derive a number of candidate speech portions (e.g. words) and (2) receiving written input to
20 disambiguate (e.g. narrow down the candidates to one result). At present with present-day speech recognition technology, this is thought less practical than the preferred embodiment. The relative merits of the two approaches depend on the relative effectiveness
25 of the two component technologies as applied to a particular pair of inputs. For example, if the user input a relatively high-quality (in terms of the system's ability to recognize) graph, and a low-quality speech-bite, the preferred approach would be more
30 effective. A system could, in fact, apply both approaches to a set of inputs, and choose the result with the higher degree of confidence.

In an alternative embodiment, then, the system according to the invention would receive spoken words
35 from the user by means of the data collection 7 and feature extraction 8 processes of Fig. 1. With

reference to Fig. 2, the extracted features would be made available to CPU 12 by data channel 27. As mentioned above in connection with the preferred embodiment, it is known in the speech-recognition art to extract these features, which are indicative typically of the intensity of sound energy in each of a number of preselected frequency bands, indicative of the overall sound intensity, and so on. In known speech recognition systems one of the most vexing problems is figuring out when one spoken word ends and the next begins, but in the apparatus of this embodiment it is assumed that the user speaks only one speech portion (word, letter, syllable, or phrase, depending on the language involved and the design choices made) at a time, in response to synthesized prompts.

CPU 12 has access to a template store located in fixed disk 16 (or alternatively, in floppy disk 17 or elsewhere) which contains all the words (or other speech portions, such as syllables, phrases, and the like) to be potentially recognized, which may range from a few hundred to as many as a hundred thousand.

The candidates in the template store are winnowed out to arrive at a relatively small number of candidates from the template store. An attempt is then made to attempt to correlate, on some appropriate measure, the data received at the touchpad with one or another of the candidates from the template store.

Measures of correlation would, in the simple case, take into account number of strokes and other relatively unambiguous aspects of the data received at the touchpad, to find a closest match with one or another of the candidates from the template store. As with the preferred embodiment, the "closest match" will generally be defined as a match that is closer by a predetermined amount than any of the other candidates,

and is objectively a match at least as close as better than some predetermined standard. In the absence of a close enough match (either because the best match was not very good, or because it was not much better than the second-place match) remedial steps are taken as described above in connection with the first embodiment.

Those skilled in the art will appreciate that while the embodiments have been disclosed with respect to one particular way of characterizing a spoken input (extraction of acoustic features and comparison with a set of acoustic-feature templates), the inventive aspects of the invention do not depend on the use of such templates. For example, one may employ instead what is known as transcription. A transcript is derived from the spoken input, where the transcript is a series of characters or symbols capable of being stored in the computer which approximate the spoken input. The transcript of the spoken input is then compared with transcripts associated with each of the candidates.

While the embodiments have been disclosed with connection to the recognition of letters, the scope of the invention should not be so limited. For example, much attention has been paid to the recognition of Chinese characters (kanji) through touchpad input means and the like. While typical ambiguity for recognition of a letter may be two or three candidates, it is not unheard-of for a given set of calligraphic data to give rise to several dozen candidate kanji. With prior art systems the inability of the algorithm to exclude all choices but one makes the entire enterprise fruitless. With the method of the invention, however, the user may also speak the kanji, and speech feature extraction and matching permit disambiguation.

In yet another application, the case of recognition of words (formed of letters) may be accomplished by the method of the invention. The user would enter a sequence of letters at the touchpad, and
5 speak the word that was spelled. What is generated in this case is a set of candidate words, allowing for ambiguity in the recognition of individual letters. Then the user speaks the word, and the candidate with the nearest match is taken to be the correct match.

10 While most of the above examples use Roman letters, it should be appreciated that the apparatus of the invention may also be set up to recognize katakana or hiragana characters, or alphabets other than the Roman alphabet, without departing from the spirit of
15 the invention. These include Chinese, Japanese, Korean, and Vietnamese, as well as editing marks, mathematical and scientific symbols, and any other graphical shape having an associated sound. Finally, depending on the language it is possible to recognize
20 entire phrases rather than words singly. In the case of phrases, an entire phrase is handwritten at the touchpad and the entire phrase is spoken by the user; the method of any of the embodiments described above is applied to the phrase just as it would be applied to a
25 letter, syllable, character or word.

Any of the above-mentioned embodiments may be improved upon by employing ranking methodologies. In the simple cases described above, one of two recognition inputs is used to derive a set of
30 candidates and the other recognition input is used to disambiguate from among the candidates. In those simple cases all the candidates on the list are treated as equally plausible in light of the recognition input that was used to generate the list; any information as
35 to whether one of the candidates was better in light of that input is discarded. In those simple cases, then,

the other of the two inputs gives rise to the correlation coefficients that permit disambiguation and selection of the "best match".

When ranking methodologies are used, a higher confidence result can be obtained. Take, for example, the embodiment in which handwriting has been recognized yielding a list of five candidate words. In the simple case, the speech recognition input (hopefully) permits the conclusion that one of the five candidates is the best match. In the "ranking" embodiment, the apparatus stores not only the list of five words resulting from the handwriting recognition, but also stores for each candidate a corresponding confidence level weighting value. Then, when the speech recognition features are compared for a match with each of the candidates, the closeness of the match is factored in together with the confidence level weighting value.

In the "ranking" embodiment, then, the results of the recognition process take into account not only how closely the spoken input agrees with one of several candidates resulting from the handwritten input, but also how "close" any given candidate is to being the right one for the handwritten input.

A computational example will show that there are several meaningful ways to accomplish the ranking. Assume a system where handwriting recognition is used to find candidate words, of which there are n , and that speech recognition is to be used to select from among the n words. Let $s(i)$ be the correlation coefficients (where i ranges from 1 to n) between the features received from the speech input and the n candidates. In the simple embodiment, if there is a successful recognition it is because for some j there is an $s(j)$ which is larger than the other $s(i)$ by at least some margin, and that $s(j)$ is itself above a predetermined

threshold; in that case candidate j is taken as the correctly recognized word.

In the ranking embodiment, another set of coefficients $h(i)$ represent the higher or lower confidence levels representing the relative likelihood that a given candidate is the correct one relative to the handwriting input. The overall conclusion as to which of the candidates is correct results from a combination of the coefficients $s(i)$ and $h(i)$. If the $s(i)$ and $h(i)$ are scaled, for example, to fall between 0 and 1, then the sum or the product of the coefficients may be used. In the case of the sum, recognition is successful if for some j there is a sum $s(j)+h(j)$ which is larger than the other $s(i)+h(i)$ by at least some margin, and that $s(j)+h(j)$ is itself above a predetermined threshold; in that case candidate j is taken as the correctly recognized word. Products may also be used.

An additional level of sophistication may be had by weighting the two inputs relative to each other; the weighting may be static or dynamic. With static weighting one of the two inputs is given a greater weight than the other; for example if the ambient noise level is always great the handwriting coefficients may be scaled to fall between 0 and 2, and then summed with the handwriting coefficients. With dynamic weighting the relative weight given to the $s(i)$ and the $h(i)$ may vary from speech portion to speech portion; if ambient noise made one particular speech portion harder to extract features from then the $s(i)$ could be given slightly less weight relative to the $h(i)$.

The ranking methodology may be applied to any of the above embodiments -- where either of the two recognition inputs is used to generate a list of

candidates and where the other of the two is used to disambiguate.

Those skilled in the art will appreciate that nothing in the invention requires that the touchpad
5 entry be first and the spoken input second; the opposite order or even simultaneous entry can be accommodated by appropriate hardware and software reconfiguration.

Claims

- 1 1. A data-entry apparatus for entering data
2 comprising handwritten word portions corresponding
3 to spoken speech portions, comprising:
4 writing-pad receiving means for receiving a
5 group of writing signals;
6 sound receiving means for detecting sounds
7 corresponding to a spoken speech portion and for
8 extracting features for said sounds;
9 template generation means for generation of a
10 multiplicity of speech portion templates, each
11 template indicative of a word portion
12 corresponding to said group of writing signals;
13 and
14 correlating means responsive to the extracted
15 features for evaluating the correlation between
16 the extracted features and the features of each
17 generated speech portion template, and for
18 identifying the generated speech portion template
19 in the multiplicity of speech portion templates
20 having the highest correlation with the extracted
21 features.
- 1 2. The data-entry apparatus of claim 1, further
2 comprising a speech synthesizer responsive to the
3 correlating means for synthesizing the speech
4 portion corresponding to the speech portion
5 template in the multiplicity of speech portion
6 templates having the highest correlation with the
7 extracted features.
- 1 3. The data-entry apparatus of claim 1 wherein the
2 writing-pad receiving means is a two-dimensional
3 touchpad.

- 1 4. The data-entry apparatus of claim 3 wherein the
2 handwritten word portions are Chinese characters
3 and the sound speech portion is a spoken Chinese
4 character.
- 1 5. The data-entry apparatus of claim 3 wherein the
2 handwritten word portions are Kanji characters and
3 the sound speech portion is a spoken Kanji
4 character.
- 1 6. The data-entry apparatus of claim 3 wherein the
2 handwritten word portions are katakana characters
3 and the sound speech portion is a spoken katakana
4 character.
- 1 7. The data-entry apparatus of claim 3 wherein the
2 handwritten word portions are hiragana characters
3 and the sound speech portion is a spoken hiragana
4 character.
- 1 8. The data-entry apparatus of claim 3 wherein the
2 handwritten word portions are words written in an
3 alphabet and the sound speech portion is a spoken
4 word.
- 1 9. A data-entry apparatus for entering data
2 comprising spoken speech portions corresponding to
3 handwritten word portions, comprising:
4 writing-pad receiving means for receiving a
5 group of writing signals;
6 sound receiving means for detecting sounds
7 corresponding to a spoken speech portion and for
8 extracting features for said sounds;
9 word-candidate generation means for
10 generation of a multiplicity of speech portion
11 word-candidates, each word-candidate indicative of

12 a candidate word portion corresponding to the
13 features for the spoken speech portion; and
14 correlating means responsive to the group of
15 writing signals for evaluating the correlation
16 between the group of writing signals and the
17 features of each generated speech portion word-
18 candidate, and for identifying the generated
19 speech portion word-candidate in the multiplicity
20 of speech portion word-candidates having the
21 highest correlation with the group of writing
22 signals.

1 10. The data-entry apparatus of claim 9, further
2 comprising a speech synthesizer responsive to the
3 correlating means for synthesizing the speech
4 portion corresponding to the speech portion word-
5 candidate in the multiplicity of speech portion
6 word- candidates having the highest correlation
7 with the group of writing signals.

1 11. The data-entry apparatus of claim 9 wherein the
2 writing-pad receiving means is a two-dimensional
3 touchpad.

1 12. The data-entry apparatus of claim 11 wherein the
2 handwritten word portions are Chinese characters
3 and the sound speech portion is a spoken Chinese
4 character.

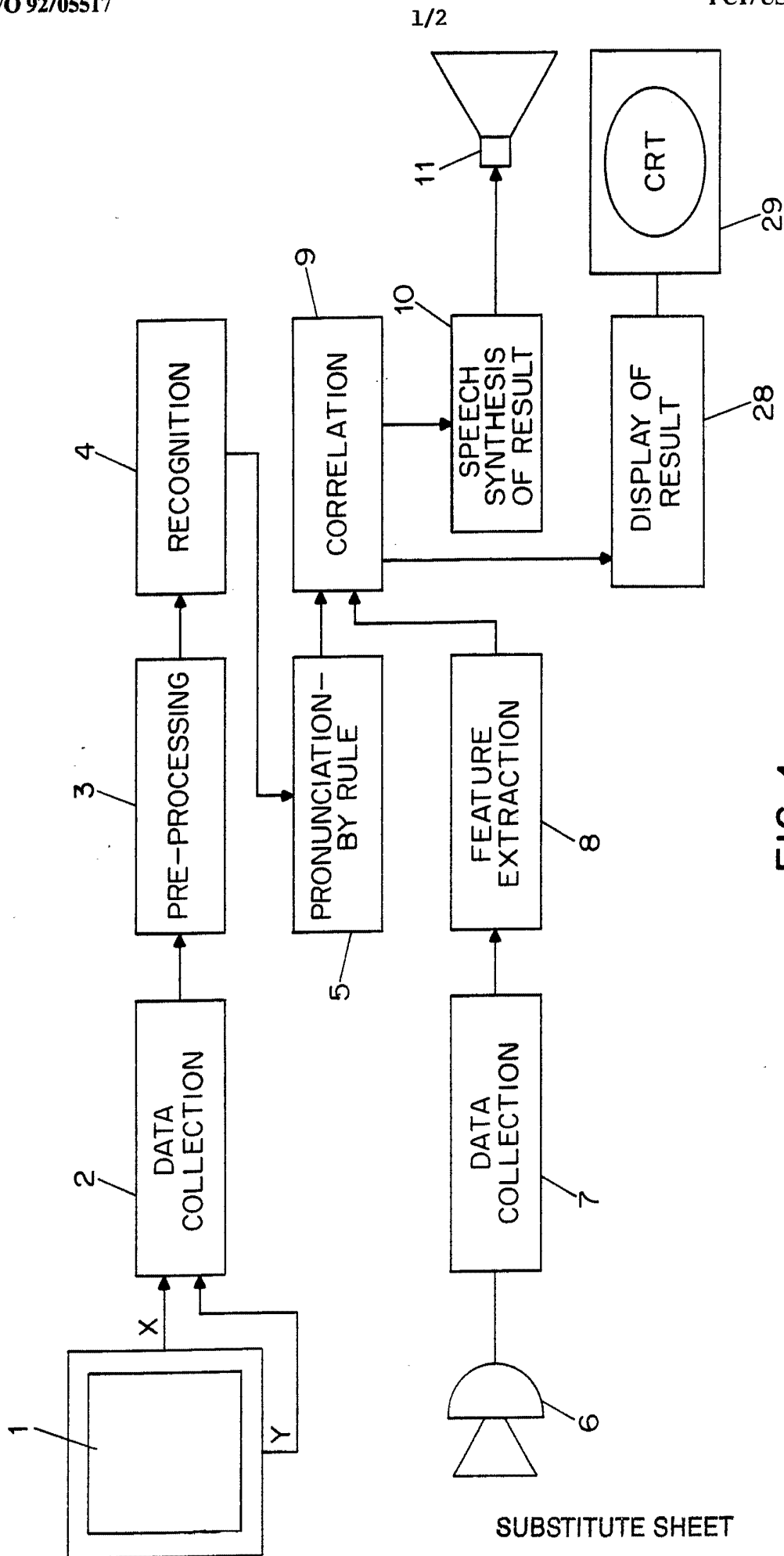
1 13. The data-entry apparatus of claim 11 wherein the
2 handwritten word portions are Kanji characters and
3 the sound speech portion is a spoken Kanji
4 character.

1 14. The data-entry apparatus of claim 11 wherein the
2 handwritten word portions are katakana characters

3 and the sound speech portion is a spoken katakana
4 character.

1 15. The data-entry apparatus of claim 11 wherein the
2 handwritten word portions are hiragana characters
3 and the sound speech portion is a spoken hiragana
4 character.

1 16. The data-entry apparatus of claim 11 wherein the
2 handwritten word portions are words written in an
3 alphabet and the sound speech portion is a spoken
4 word.



SUBSTITUTE SHEET

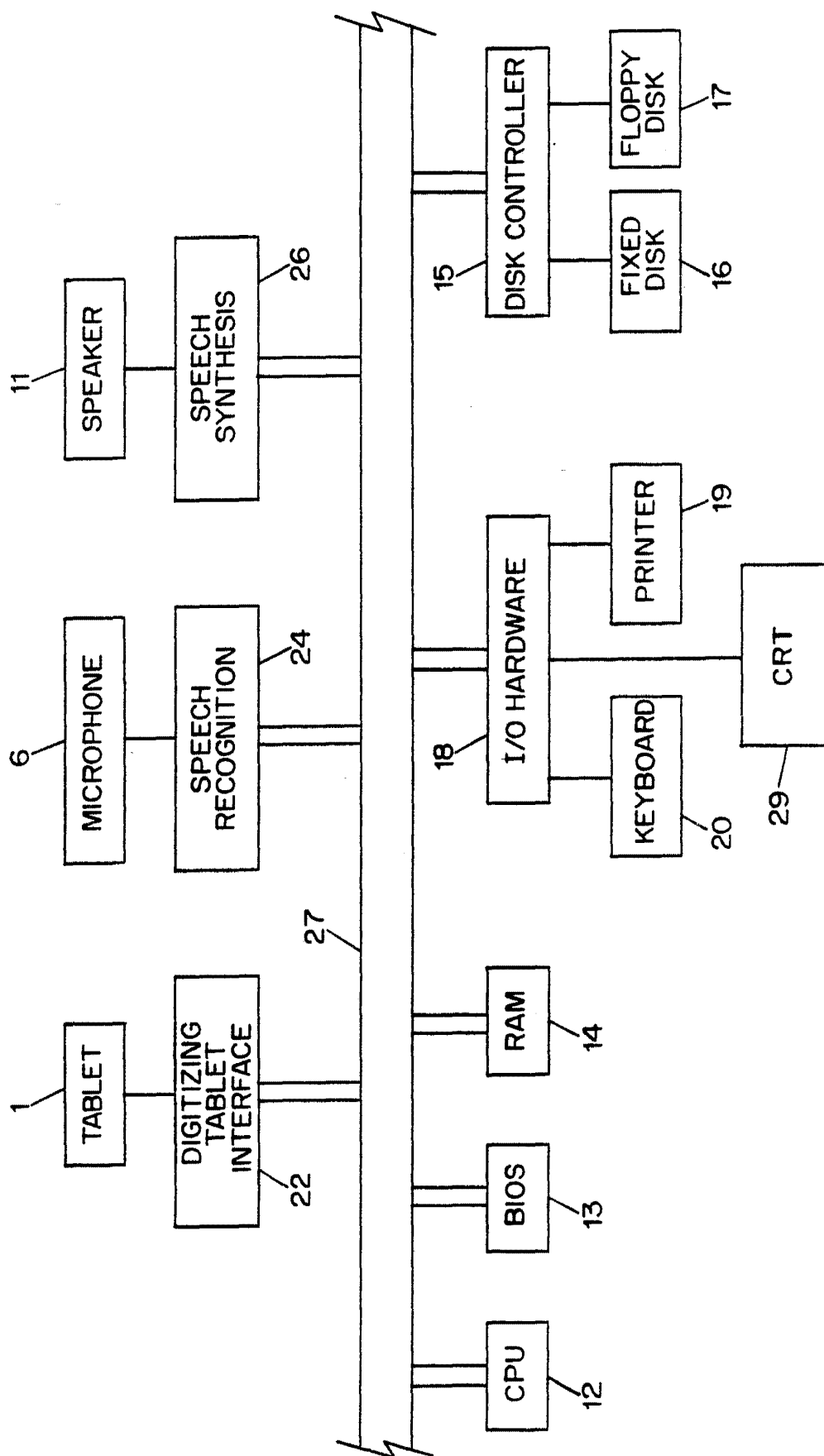


FIG. 2

SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

PCT/US 91/06874

International Application No

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

Int.Cl. 5 G06K9/22; G06F3/16

II. FIELDS SEARCHED

Minimum Documentation Searched⁷

Classification System	Classification Symbols
Int.Cl. 5	G06K ; G06F ; G10L

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched⁸III. DOCUMENTS CONSIDERED TO BE RELEVANT⁹

Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
X	PATENT ABSTRACTS OF JAPAN vol. 7, no. 249 (P-234)5 November 1983 & JP,A,58 134 371 (NIPPON DENKI K.K.) 10 August 1983 see abstract ---	1,3,9,11
X	PATENT ABSTRACTS OF JAPAN vol. 10, no. 45 (P-430)21 February 1986 & JP,A,60 189 070 (FUJITSU K.K.) 26 September 1985 see abstract ---	9
A	PATENT ABSTRACTS OF JAPAN vol. 11, no. 87 (P-557)17 March 1987 & JP,A,61 240 361 (HITACHI ELECTRONICS ENG. CO. LTD.) 25 October 1986 see abstract --- -/--	2,10

¹⁰ Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

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

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

"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

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
22 JANUARY 1992	28. 01. 92
International Searching Authority	Signature of Authorized Officer
EUROPEAN PATENT OFFICE	Alfredo Prem

Form PCT/ISA/210 (second sheet) (January 1985)

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.
A	PATENT ABSTRACTS OF JAPAN vol. 7, no. 146 (P-206)25 June 1983 & JP,A,58 058 637 (NIPPON DENKI K.K.) 7 April 1983 see abstract ---	1,4,6,7, 12,14,15
A	PATENT ABSTRACTS OF JAPAN vol. 6, no. 137 (P-130)24 July 1982 & JP,A,57 059 241 (SASAKI NOBUHIKO) 9 April 1982 see abstract ---	1,5,13