

Veröffentlichung im Amtsblatt	<input checked="" type="checkbox"/> Ja/Nein
Publication in the Official Journal	<input checked="" type="checkbox"/> Yes/No
Publication au Journal Officiel	<input checked="" type="checkbox"/> Oui/Non

Aktenzeichen / Case Number / N^o du recours : T 458/89 - 3.4.1

Anmeldenummer / Filing No / N^o de la demande : 85 303 666.3

Veröffentlichungs-Nr. / Publication No / N^o de la publication : 0 164 945

Bezeichnung der Erfindung: Methods of and apparatus for speech recognition
Title of invention:
Titre de l'invention :

Klassifikation / Classification / Classement : G10L 5/06

ENTSCHEIDUNG / DECISION

vom / of / du 8 August 1990

Anmelder / Applicant / Demandeur : Sony Corporation

Patentinhaber / Proprietor of the patent /
Titulaire du brevet :

Einsprechender / Opponent / Opposant :

Stichwort / Headword / Référence :

EPÜ / EPC / CBE Art. 56

Schlagwort / Keyword / Mot clé : "inventive step (yes)"

Leitsatz / Headnote / Sommaire



Case Number : T 458/89 - 3.4.1

D E C I S I O N
of the Technical Board of Appeal 3.4.1
of 8 August 1990

Appellant : Sony Corporation
7-35 Kitashinagawa 6-Chome Shinagawa-ku
Tokyo 141 (JP)

Representative : Thomas, Christopher Hugo et.al.
D. Young & Co.
10 Staple Inn
London WC1V 7RD (GB)

Decision under appeal : Decision of Examining Division 037 of the
European Patent Office dated 1 March 1989
refusing European patent application
No. 85 303 666.3 pursuant to Article 97(1) EPC

Composition of the Board :

Chairman : K. Lederer
Members : H. Reich
L. Mancini

Summary of Facts and Submissions

- I. European patent application No. 85 303 666.3 (publication number 0 164 945) was refused by decision of the Examining Division.
- II. The reason given for the refusal was that the subject-matter of independent Claims 1 and 12, received 7 June 1988, did not satisfy the requirements of Articles 52 and 56 EPC having regard to documents:

D1: "IEEE Transactions on Computers", Vol. C-20, No. 9, 9/71, pages 972-978;

D2: "International Journal of Man-Machine Studies", Vol. 14, 1981, pages 13-27; and

D3: DE-A-2 448 908.

The Examining Division took the view that a skilled person would arrive at the subject-matter in particular of Claim 1 by combining in an obvious way the characteristics of the pre-characterising portion of Claim 1 which are known from document D1, with a sampling step of a plotted trajectory such as known from document D3, and measures known from document D2 in which the lengths of plotted and registered trajectories are determined and used for comparison. This combination would not produce any unexpected results and thus not involve an inventive step.

- III. The Appellant lodged an appeal against the decision. In the Grounds of Appeal, the Appellant argued inter alia, that there would be a very substantial distinction between Claim 1 and the disclosure of document D3. In the method known from document D3 the sampling of the plotted

trajectory would not be effected "at a predetermined length therealong" but the distances between the points in the final curve would be totally arbitrary and result in a more complex matching step.

- IV. In response to a communication of the Board of Appeal, inviting the Appellant to further amend his claims, in particular in order to make Claims 1 and 12 include the essential features of the invention - concerning the particular parameter space and a constant sampling interval - as required by Article 84 EPC, the Appellant now requests that the contested decision be set aside and that a patent be granted on the basis of the following documents:

Claims: 1-22, received 29 June 1990 with letter dated 25 June 1990;

Description: pages 1-3, 8, 10, 17, received 29 June 1990 with letter dated 25 June 1990; and pages 4-7, 9, 11-16, 18-21 according to EP-A-0 164 945;

Drawings: pages 1/8 to 8/8 according to EP-A-0 164 945.

- V. Independent Claims 1 and 12 read as follows:

"1. A method of speech recognition comprising the steps of:
segmenting an input speech signal;
deriving a plurality of time-sequential acoustic parameters from the periodically sampled energies of the segmented speech signal in a predetermined number of neighbouring frequency bands;

plotting a trajectory in an acoustic parameter space on the basis of dots, the co-ordinates of each dot representing the values of the acoustic parameters within the same sampling period, to form a time-normalized trajectory of said segmented speech signal;
providing a plurality of corresponding registered standard trajectories, each standard trajectory representing one segmented speech signal;
matching said plotted trajectory with said registered standard trajectories;
and
producing an indication of the results of the matching; characterized by:
the co-ordinates of each dot in the acoustic parameter space corresponding to the values of the energy of the speech signal in each of said neighbouring frequency bands; and
sampling said plotted trajectory at a predetermined constant sampling interval therealong, thereby producing new speech recognition data replacing said plotted trajectory in the matching step with said registered standard trajectories.

12. Apparatus for speech recognition, the apparatus comprising:

means (17) for segmenting an input speech signal;
means (2) for producing a plurality of time-sequential acoustic parameters ($P_i(n)$) from the periodically sampled energies of the segmented speech signal in a predetermined number of neighbouring frequency bands;
means (22,23,24) for producing a plurality of speech recognition data from said plurality of time-sequential acoustic parameters;
means (4) for registering a plurality of standard speech recognition data;

means (25,6) for matching said produced speech recognition data with previously registered standard speech recognition data; and
output means (6) for producing an output representing the results of the matching processing in said means (25,6) for matching;
characterized in that:
said speech recognition data producing means (22,23,24) comprises:
first calculating means (22) receiving said time-sequential acoustic parameters for calculating in an I-dimensional parameter space, I being the number of neighbouring frequency bands, in which space the coordinates of each point correspond to the values of the energy of the speech signal in each of said frequency bands within the same sampling period, a distance $S(n)$ between acoustic parameters produced at a sampling time (n) of the segmented speech signal and respective acoustic parameters produced at adjacent sampling times (n+1) and for calculating a trajectory length (SL) in said parameter space between the acoustic parameters of the first sampling time and the acoustic parameters of the last sampling time by summing all of said intermediate calculated distances, second calculating means (23) for determining a constant sampling interval (T) by dividing said trajectory length (SL) in said parameter space into a predetermined number of equally large portions, and third calculating means (24) receiving said constant sampling interval (T) and said acoustic parameters ($P_i(n)$) for calculating said speech recognition data (Q_i) by sampling the trajectory of the acoustic parameters in the parameter space at said constant sampling interval."

Claims 2 to 11 are dependent on Claim 1, and Claims 13-22 are dependent on Claim 12.

Reasons for the Decision

1. The appeal is admissible.
2. The subject-matter of the current set of claims does not extend beyond the content of the application as originally filed. The subject-matter of Claim 1 comprises that of originally filed Claims 1 and 10 and the subject-matter of Claim 12 corresponds to that of originally filed Claims 14 and 15. Both said claims contain in addition the following features: the specification of the parameter space, i.e. that the co-ordinates of each dot (point) correspond to the values "of the energy of the speech signal in each of said neighbouring frequency bands" as disclosed in the originally filed description, page 7, lines 10 to 17, and the fact that the plotted trajectory is sampled "at a predetermined constant sampling interval" according to the original description, page 8, lines 29 to 31.

The feature in appealed Claim 1 concerning the use of the length of a plotted trajectory in the comparison with registered data as additional matching parameter has been transferred into present Claim 11, because it is only part of the particular embodiment of the invention according to Fig. 10 (see length data adder 31), and is absent in the embodiments of Figs. 1, 7 and 8. The feature in appealed Claim 12 concerning the means for correcting the matching distance between sampled and registered data has been transferred into present Claim 13, being comprised only in the embodiment of Fig. 11.

Dependent Claims 2 to 9 correspond to original Claims 2 to 9, Claim 10 to original Claim 11, Claim 11 to original Claims 12 and 13, Claim 13 to original Claim 26, and Claims 14 to 22 to original Claims 16 to 24. The subject-matter of Claim 20 has been clarified on the basis of Figs. 10 and 11, and that of Claim 21 on the basis of Fig. 7.

Furthermore claims and description have been amended in order to satisfy Rule 35(13) EPC. The description has been adapted with regard to Rules 27(1)(c) and 27(1)(d) EPC.

There is, therefore, no objection under Article 123(2) EPC to the current set of application documents.

3. Novelty

- 3.1 None of the documents cited in the European Search Report describes a method or an apparatus for speech recognition, wherein the simultaneously sampled energies of a segmented speech signal in each of the predetermined neighbouring frequency bands are used as co-ordinates for plotting a trajectory (measured acoustic parameters) - each dot of the trajectory resulting from the subsequent periodical sample - and wherein this experimentally determined primary trajectory is again sampled at a constant sampling interval to produce a calculated, interpolated secondary trajectory in the same co-ordinate space (sampled speech recognition data) which is then matched with registered standard trajectories.
- 3.2 The method and apparatus known from document D1 correspond to the features in the pre-characterising parts of Claims 1 and 12 respectively. In this prior art, the trajectory is plotted in a space of co-ordinates, the number of which is reduced by a principle component

analysis. The outputs of all (17) neighbouring frequency bands are measured, but only 3 outputs, having maximal amplitudes, are selected for plotting the trajectory in a space of only 3 principle dimensions. This measured three-dimensional trajectory is directly - without further sampling in order to form a secondary trajectory - matched with standard trajectories registered also in the corresponding three-dimensional space.

3.3 In the method and apparatus for speech recognition according to document D3, the co-ordinate space is further reduced to 2 dimensions, corresponding to the two formants (i.e. centres of gravity in the frequency spectrum of a speech signal, which centres are independent from the tone pitch) with the lowest frequencies of a phoneme (smallest unit of significant sound). In this space a trajectory is plotted according to the formant frequencies for a complete syllable (segmented speech signal). This known primary trajectory is not re-sampled with constant sampling intervals but simplified according to a different principle: For all vowels of the syllable one average value (centre of gravity) of the formant frequencies is formed, and for the consonants of the syllable the trajectory is linearised by thinning it to the 4 points with the smallest angles formed by the connecting lines to neighbouring points; see Fig. 13 to 16 of D3. The centre of gravity of the vowel co-ordinates and the linearised consonant trajectory are then matched with corresponding registered values.

3.4 In the method and apparatus of document D2 a phoneme is characterised by its vocal characteristics and its duration. The vocal characteristics indicate whether a sound is Fricative, Voiced or Silent. These three categories are subdivided by parameters which, however, are not further specified in document D2.

3.5 The remaining document mentioned in the European Search Report, IBM-Technical Disclosure Bulletin, Vol. 25, No. 11A, April 1983, pages 3448-3452, used the utterance duration and the energy envelope of the utterance for matching with registered data. The energy envelope is formed by the sum of the energies in all frequency bands (filters) at uniform sampling intervals of the utterance.

3.6 Thus, the subject-matter of Claims 1 and 12 is considered novel in the sense of Article 54 EPC.

4. Inventive Step

4.1 Starting from the nearest prior art as disclosed in document D1, the objective problem underlying the present invention as claimed in Claim 1 is to remove the influence of speech utterance speed fluctuations on the matching step of measured and registered data and to obtain, even when the speech fluctuates, a relatively high recognition ratio; see the description, page 7, line 23 to page 8, line 8 and page 12, lines 13 to 23. It should be noted that the "time normalisation" applied in the method of document D1 is different from the above problem. The known time normalisation stretches or compresses linearly a speech utterance as a whole. Thus, the trajectory is linearly stretched or compressed in total whereby individual constant speech speeds of a complete utterance are normalised for the matching step. Whereas the present invention eliminates speech speed fluctuations within an utterance, i.e. within the trajectory.

4.2 The problem underlying the present invention is solved by the sampling step defined in the characterising part of Claim 1, or by the first to third calculating means claimed in Claim 12, i.e. by resampling the experimentally

obtained trajectory along its length at a predetermined constant sampling interval.

- 4.3 The cited prior art - and in particular document D3 - gives no hint to the skilled person to eliminate the influence of speech utterance speed fluctuation on speech recognition by taking samples from the experimental trajectory at constant intervals, for the following reasons:

Only in the method known from document D3 is the experimental trajectory reshaped before the matching step. Apart from the diverging co-ordinate space of the trajectory in document D3, the known reshaping steps of the trajectory as such have nothing in common with the resampling technique applied in Claims 1 and 12 of the present invention. The known reshaping process according to document D3 uses a different simplification technique for those trajectory parts which belong to vowels, and for those trajectory parts which belong to the consonants of the syllable (segmented speech signal). It averages the co-ordinates of the time-sequential samples for vowels to form a centre of gravity and linearises the trajectory of the time sequential samples for consonants by eliminating the samples with little influence on the characteristic shape of the trajectory, i.e. those samples (dots) forming large angles with neighbouring samples. There is no technical equivalence between the formation of the centre of gravity of a variety of samples or the linearisation of a trajectory of samples on the one hand and a re-sampling of a sampled trajectory on the other hand.

Moreover the reshaping of the trajectory according to document D3 serves a different purpose: It simplifies the necessary circuitry of the hardware and reduces the memory capacity.

- 4.4 Hence, in the Board's view a skilled person is not able to derive from the prior art that re-sampling an experimental trajectory as claimed in Claim 1 eliminates speech utterance speed variation within the segmented speech signal from the matching step. Accordingly, it is not obvious for a skilled person to provide "third calculating means" for said sampling and "first and second calculating means" for determining said constant sampling interval as claimed in Claim 12.
- 4.5 Thus, the subject-matter of Claims 1 and 12 is considered to involve an inventive step in the sense of Article 56 EPC.
5. Hence, it follows that Claims 1 and 12 are allowable.
6. Therefore, Claims 2 to 11, depending on Claim 1, and Claims 13 to 22, depending on Claim 12, are also allowable.

Order

For these reasons, it is decided that:

1. The decision under appeal is set aside.
2. The case is remitted to the first instance with the order to grant a European patent on the basis of the following documents:

Claims: 1-22, received 29 June 1990;

Description: pages 1-3, 8, 10, 17, received 29 June 1990,
pages 4-7, 9, 11-16, 18-21 according to EP-A-
0 164 945;

Drawings: pages 1/8 to 8/8 according to EP-A-
0 164 945.

The Registrar:

The Chairman:

P. Martorana

K. Lederer