BESCHWERDEKAMMERN DES EUROPÄISCHEN PATENTAMTS

BOARDS OF APPEAL OF THE EUROPEAN PATENT OFFICE

CHAMBRES DE RECOURS DE L'OFFICE EUROPEEN DES BREVETS

Internal distribution code:
(A) [ ] Publication in OJ
(B) [ ] To Chairmen and Members
(C) [ ] To Chairmen
(D) [X] No distribution

Datasheet for the decision of 30 April 2013

Case Number: T 1555/08 - 3.4.01

Application Number: 03769735.6

Publication Number: 1565906

IPC: G10L 15/26

Language of the proceedings: EN

Title of invention:
Speech recognition device and method

Applicant:
Nuance Communications Austria GmbH

Headword:
-

Relevant legal provisions (EPC 1973):
EPC Art. 56

Keyword:
"Juxtaposition of known means"

Decisions cited:
T 0991/07, T 1704/06, T 0823/04

Catchword:
-

Europäisches Patentamt | European Patent Office | Office européen des brevets

Beschwerdekammern | Boards of Appeal | Chambres de recours

Case Number: T 1555/08 - 3.4.01

D E C I S I O N
of the Technical Board of Appeal 3.4.01
of 30 April 2013

Appellant: (Applicant)
Nuance Communications Austria GmbH
Triester Straße 64
A-1101 Wien (AT)

Representative:
Driver, Virginia Rozanne
Page White & Farrer
Bedford House
John Street
London WC1N 2BF (GB)

Decision under appeal:
Decision of the Examining Division of the European Patent Office posted 7 March 2008 refusing European patent application No. 03769735.6 pursuant to Article 97(2) EPC.

Composition of the Board:

Chairman: F. Neumann
Members: P. Fontenay
C. Schmidt

Summary of Facts and Submissions

I. The appeal lies from the decision of the examining division to refuse European patent application No. 03 769 735.6. The decision was dispatched on 7 March 2008.

The decision relied, primarily, on the finding that the subject-matter of independent claim 1 of the main request and first and second auxiliary requests did not involve an inventive step in the sense of Article 56 EPC 1973. A similar finding applied to the independent claims of the main and auxiliary requests directed to the corresponding method and computer program product. In this respect, the examining division held that the claimed invention, as defined in claim 1 of the main and first auxiliary requests, was a mere juxtaposition of known independent features in the field of speech recognition. Particular reference was made to document US-A-2002/0138272 (D2) and to various additional aspects regarding speech recognition as disclosed in documents US-A-2002/0087306 (D1) and US-A-2002/0087311 (D3).

II. The appellant (applicant) filed an appeal against the above decision by notice of appeal received on 27 March 2008. The prescribed appeal fee was paid on the same day. The written statement setting out the grounds of appeal was received on 10 July 2008. It was requested that the decision under appeal be set aside and a patent be granted on the basis of various sets of claims according to a main request or auxiliary requests 1 to 5.

III. In the statement of grounds, the appellant presented arguments which, in its opinion, established that the claimed invention involved an inventive step in view of document WO-A-96/08215 (D4), which it considered to represent the most pertinent art. In the appellant's view, the claimed invention permitted to improve the accuracy of speech recognition and to achieve a considerably more exact modelling of the language during speech recognition.

Referring more particularly to the other prior art documents D1, D2 and D3, the appellant stressed that no motivation could be found in any of these documents to provide for the claimed measures.

IV. On 4 January 2013, summons to attend oral proceedings were issued.

In a communication of the Board pursuant to Article 15(1) of the Rules of Procedure of the Boards of Appeal (RPBA) annexed to the summons, the Board informed the appellant of its provisional assessment of the requests then on file.

In particular, the Board observed that it could not identify any difference between the subject-matter of the independent claims of the main and first auxiliary request and the teaching of document D4. While acknowledging that the second auxiliary request defined new subject-matter, the Board expressed doubts as to its inventive merits. Moreover, in the Board's view, the additional features recited in the independent claims of auxiliary requests 3 to 5 were already implicitly present in the claims of the main request

and auxiliary request 1 and 2, respectively, so that the conclusion to be reached with regard to the main request and auxiliary request 1 and 2 were also to apply to auxiliary request 3 and 4, respectively.

Although focusing on the teaching of document D4, the Board further indicated that it did not find fault in the analysis carried out by the examining division according to which the claimed speech recognition device and method resulted, in essence, from mere juxtaposition of known independent means and techniques well known in the field of speech recognition.

V. With letter of reply dated 28 March 2013, the appellant filed a set of claims 1 to 3 according to a new main request and a single claim 1 according to a new first auxiliary request. The new requests replaced all previous requests filed with the statement of grounds. The claims of the new requests had been amended in order to take into account the comments made by the Board in its provisional opinion. In the accompanying letter, the appellant presented arguments which, in its view, established that the claimed speech recognition device according to the new requests did indeed involve an inventive step.

VI. In view of the substantial amendments made to the claims of the new requests and in particular the introduction in the independent claims of the features relating to the fact that the speech information was processed in parts comprising a plurality of frames, the representative was informed by facsimile on 26 April 2013 that reference would be made, if necessary, during the oral proceedings to document

EP-A-1 160 768 (D5) cited in the International Search report.

VII. Oral proceedings before the Board took place on 30 April 2013. As had previously been announced in a communication of 26 April 2013, the appellant was not represented.

VIII. Claim 1 of the appellant's main request reads:

"*1. A speech recognition device comprising:*

speech information receiving means (2) for receiving speech information (SI) from a plurality of participants and capable of being received via at least two reception channels;

reception-channel recognition means (18) to recognize a reception channel used to receive the speech information (SI) and to generate channel information (CHI) representing the reception channel that is recognized;

feature vector (FV) extraction means for generating and emitting feature vectors by taking into account the channel information (CHI);

first language-property recognition means (20) to perform acoustic segmentation to provide segmentation information (ASI) by using the feature vectors (FV) and continuously taking in to [sic!] account the channel information (CHI);

second language-property recognition means (21) to determine language information (LI) indicating the language of the speech information (SI) by using the feature vectors (FV) and continuously taking in to [sic!] account the channel information (CHI) and the segmentation information (ASI); and

speech recognition means (24) arranged to recognize text information (TI) corresponding to the speech information (SI) received using the segmentation and the language information;

wherein the received speech information is processed in parts, with each part comprising a plurality of frames;

and wherein the reception-channel recognition means recognizes the reception channel on a frame by frame basis of the speech information (SI) and continuously updates the channel information (CHI);

and wherein the first language-property recognition means determines segmentation information (ASI) for each frame of the speech information;

and wherein the second language-property recognition means determines language information (LI) indicating the language of each frame of the speech information using the segmentation information for the respective frame of the speech information;

and wherein the speech recognition means recognizes the text information (TI) for each frame by continuously taking into account the segmentation information and the language information indicating the language of the speech information (SI) for the respective frame of the speech information."

Claims 2 and 3 of the main request are dependent claims.

Claim 1 of the first auxiliary request results from a combination of claims 1, 2 and 3 of the main request. It reads:

"*1. A speech recognition device comprising:*

speech information receiving means (2) for receiving speech information (SI) from a plurality of participants and capable of being received via at least two reception channels;

reception-channel recognition means (18) to recognize a reception channel used to receive the speech information (SI) and to generate channel information (CHI) representing the reception channel that is recognized;

feature vector (FV) extraction means for generating and emitting feature vectors by taking into account the channel information (CHI);

first language-property recognition means (20) to perform acoustic segmentation to provide segmentation information (ASI) by using the feature vectors (FV) and continuously taking in to [sic!] account the channel information (CHI);

second language-property recognition means (21) to determine language information (LI) indicating the language of the speech information (SI) by using the feature vectors (FV) and continuously taking in to [sic!] account the channel information (CHI) and the segmentation information (ASI); and

speech recognition means (24) arranged to recognize text information (TI) corresponding to the speech information (SI) received using the segmentation and the language information;

wherein the received speech information is processed in parts, with each part comprising a plurality of frames;

and wherein the reception-channel recognition means recognizes the reception channel on a frame by

frame basis of the speech information (SI) and continuously updates the channel information (CHI);

and wherein the first language-property recognition means determines segmentation information (ASI) for each frame of the speech information;

and wherein the second language-property recognition means determines language information (LI) indicating the language of each frame of the speech information using the segmentation information for the respective frame of the speech information;

the speech recognition device further comprising third language-property recognition means (22) to recognize speaker group information (SGI) representing a recognized speaker group, wherein the third language-property recognition means uses the feature vectors (FV) and continuously takes into account the segmentation information (ASI) and language information (LI) and channel information (CHI) to generate the speaker group information representing the speaker group recognized for the respective speech information (SI);

the speech recognition device further comprising fourth 1anguage-property recognition means (23) to recognize the context of the speech information, wherein the fourth 1anguage-property recognition means uses the feature vectors (FV) and continuously takes into account the segmentation information (ASI), language information (LI) and speaker group information (SGI) and channel information (CHI) to generate context information (CI) representing the context for the respective speech information (SI);

and wherein the speech recognition means recognizes the text information (TI) for each frame of the speech information by continuously taking into account the segmentation information (ASI), the

language information (LI) indicating the language of the respective frame of the speech information (SI), the speaker group information (SGI) representing the speaker group recognized for the respective frame of the speech information (SI), and the context information (CI) representing the context of the respective frame of the speech information (SI)."

Reasons for the Decision

1. *Applicable law*

This decision is issued after the entry into force of the EPC 2000 on 13 December 2007 whereas the present application was filed before this date. Reference is therefore made to the relevant transitional provisions indicating which articles and rules of the EPC 1973 and the EPC 2000 are applicable to the present application. References to articles or rules of the old EPC are followed by the indication "1973" (cf. EPC, Citation practice).

2. *Admissibility of the appeal*

The notice of appeal and the statement of grounds comply with the requirements of Articles 106 to 108 EPC and Rule 99 EPC. The appeal is, therefore, admissible.

3. *Admissibility of the new main request and auxiliary request*

Although filed late, it is first noted that the new main request and auxiliary request appear to address

various issues raised by the Board in its preliminary opinion. It is further observed that the number of requests has been substantially reduced compared with the number of requests filed with the statement of grounds and that the differences between the new main request and auxiliary request are straightforward and directly identifiable. The filing of the new requests is therefore clearly beneficial to the economy of the procedure. Hence, the Board - exercising its discretional power under Article 13(1) RPBA - decides to admit the new main request and auxiliary request filed with letter of 28 March 2013 into the appeal proceedings.

4. *Further procedural matters*

The Board takes due account of the fact that the analysis relied upon by the Board in its provisional opinion regarding the lack of novelty of the subject-matter of claim 1 of the main request and auxiliary requests 1, 3 and 4 then pending no longer applies. However, the substantial amendments which have been made in independent claim 1 of the main request and auxiliary request and, in particular, the incorporation in these independent claims of the new features relating to the fact that the speech information is processed in parts comprising a plurality of frames requires a completely new assessment of the case. The circumstance that the appellant abstained from participating in the oral proceedings does not prevent the Board deciding on the case and basing its decision on objections which are, at least partly, new to the appellant.

This approach is in conformity with established case law of the boards of appeal (cf. e.g. T 991/07, point 2; T 1704/06, point 7; T 823/04, point 1, none of them published) and Article 15(3) RPBA (former Article 11(3) RPBA), which sets out that the Board shall not be obliged to delay any step of the proceedings, including its decision, by reason only of the absence at the oral proceedings of any party duly summoned. In this respect, the Board shares the opinion expressed in the explanatory note to this Article, as it appears in the document addressed to the Administrative Council of the European Patent Organisation regarding amendments to the Rules of Procedure of the Boards of Appeal (RPBA) and which reads: "... *This provision does not contradict the principle of the right to be heard pursuant to Article 113(1) EPC since that Article only affords the opportunity to be heard and, by absenting itself from the oral proceedings, a party gives up that opportunity...*" (cf. CA/133/02, 12 November 2002).

5. *Main request (Inventive step - Article 56 EPC 1973)*

5.1 In essence, the Board concurs with the examining division in its finding that the claimed invention results from the mere juxtaposition of known entities, each fulfilling its own functionality. In the absence of any effect extending beyond the sum of effects achieved by each unit constituting the claimed device, no inventive step can be recognised in the claimed association of known functional units (cf. section 5.2 below). Moreover, the details required to implement the claimed speech recognition device do not appear to involve any skills extending beyond what may be expected from the skilled person, at least insofar as

the advantages of certain measures are straightforward (cf. section 5.3 below). For these reasons the subject-matter of claim 1 of the main request cannot be considered to involve an inventive step.

5.2 It is firstly observed that a speech recognition device necessarily comprises speech information receiving means. The possibility for such a device to receive information from a plurality of participants via at least two reception channels is implicitly disclosed in document D2 where such devices are to be used in network services (cf. D2, paragraphs [0013] and [0014]). The necessity to recognize the reception channel used to receive the speech information and to generate corresponding channel information is explicitly acknowledged in D2 (cf. D2, paragraph [0017], first sentence; [0021]).

Feature vector extraction means for generating and emitting feature vectors are also well-known, as such, and are rendered necessary by the need to identify parameters (features) illustrative of the speech information to be recognised which are required for further processing of the speech information. Such means are, for example, also disclosed in D2 where these features or parameters are required to identify speech recogniser configuration parameters. In D2, these means take also into account the channel information (cf. D2, paragraphs [0015] to [0017]), as recited in claim 1 of the main request.

First language-property recognition means to perform acoustic segmentation to provide segmentation information are inherent to speech recognition

techniques and are, for example, disclosed in document D4 (cf. page 1, lines 10-18; page 17, lines 16-20). While in D4, this first language-property information is already sufficient to identify the phonetic dictionary which permits words to be associated to the phoneme code sequences, document D2, in contrast, puts particular emphasis on the need to identify additional language properties to perform this operation. It is assumed, in this respect, that the dictionary to be used in D4 is already known beforehand. In a different environment with users of different genders, native language, accents etc. (cf. D2, paragraph [0015]) making use of various channels, as in the present invention, the use of second language-property recognition means appears indispensable. Indeed such means are known from D2 and are used to identify a speech model and speech recognizer configuration parameters tailored to the user (cf. D2, paragraph [0015], [0017]; [0021]). In the case where different language are spoken by different users then the language of the speech information will obviously have to be determined.

Speech recognition means arranged to recognize text information on the basis of various properties previously identified during the processing are also inherent to speech recognition means and reflect the very purpose of such speech recognition devices (cf. D1, D2, D3 and D4). While in document D4, in which the phonetic dictionary appears to be pre-established, it would appear to be sufficient to base the speech recognition solely on the acoustic segmentation. Document D2 makes it clear that, in an environment with a large number of users making use of various channels,

the quality of the recognised text information depends on the identification of additional speech properties (cf. D2, paragraph [0015]). In particular, the use of language information would be indispensable in an environment where user language may change.

In the Board's judgement, the claimed subject-matter results thus from the integration of various means, each of which is, *per se*, inherent to speech recognition systems together with additional recognition means, each being intended to improve the accuracy of recognition. In any case, the Board is unable to identify any technical effect achieved by the claimed device which goes beyond the effect which would be expected by the aggregation of such claimed means, each means fulfilling their proper function.

Consequently, no inventive merit can be seen in the aggregation of the various functionalities referred to above. The same applies to the integration in one single system of the corresponding means.

5.3 Despite this finding, the Board has, additionally, to decide whether the inter-relationships which exist between the various units and the manner the speech data are processed could nevertheless justify the existence of an inventive step.

It is, firstly, observed that the specific sequence of the various processing steps required to carry out speech recognition is not in fact something which may be selected by the skilled person but, instead, is determined by the very purpose of each processing step in the whole recognition process. In an environment

with multiple reception channels, the identification of the reception channel actually used at the time speech information is received constitutes, in this regard, a prerequisite for further speech processing. Segmentation is only possible if information concerning both channel and characteristics ("feature vectors") of the signal received are already available. Similarly, the determination of language information indicating the language of the speech information is only effective if segmentation information has been previously made available. Finally, the text information corresponding to the speech information can most efficiently be determined once the segmentation information and language information have been determined.

Secondly, the processing of the speech information in parts, with each part comprising a plurality of frames, is known as such from document D5 (cf. [0002], [0009]). It is stressed, in this respect, that the concept of frame in document D5 corresponds to what is described in the present invention as a frame, i.e. a speech signal for a period of about 10 ms (cf. D5, [0002], [0010]).

In the Board's judgement, it would also be obvious in a dynamic environment, in which the reception channel as well as the language and possibly other parameters as to the speaker may change, to adapt the system so that it reacts to such a changing environment. More concretely, this implies that each functional unit of the claimed device would process the data on the basis of the smallest unit of data actually available and that it would rely, to do so, on the properties of the

signal already identified at that particular stage of the processing.

In summary, the development of a device as claimed in claim 1 requires solely the application of standard measures in a straightforward manner. The subject-matter of claim 1 of the main request is therefore not inventive within the meaning of Article 56 EPC 1973.

6. *Auxiliary request (Inventive step - Article 56 EPC 1973)*

A similar finding applies to the independent claim of the auxiliary request. It is stressed, in this respect, that it is known that the performance of a speech recognition device is improved by taking into account additional parameters such as the speaker group (cf. D2, 0015]) or the context of the speech information (cf. D4, page 15, line 19 - page 16, line 26; page 19, lines 3-23). While it is acknowledged that the additional means recited in claim 1 of the auxiliary request might indeed permit a considerably more exact modelling of the language to be achieved, the Board is unable to identify any technical effect which would extend beyond what may be expected from the use of such additional processing means.

Moreover, as set out above with regard to the main request, it would be obvious in a dynamically changing environment to identify the speaker group and the context of the speech information by relying on the various features which have been gathered for each particular frame of data so as to better cope with such a changing environment.

Consequently, the subject-matter of the sole claim of the auxiliary request does not involve an inventive step within the meaning of Article 56 EPC 1973 either.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar The Chairman

D. Meyfarth F. Neumann