

Internal distribution code:

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

**Datasheet for the decision
of 5 October 2021**

Case Number: T 1898/17 - 3.4.01

Application Number: 06734403.6

Publication Number: 1851758

IPC: G10L15/22, G10L15/06

Language of the proceedings: EN

Title of invention:

METHODS AND SYSTEMS FOR CONSIDERING INFORMATION ABOUT AN
EXPECTED RESPONSE WHEN PERFORMING SPEECH RECOGNITION

Applicant:

Vocollect, Inc.

Headword:

Speech recognition / Vocollect

Relevant legal provisions:

EPC Art. 56, 54

RPBA 2020 Art. 13(1), 13(2)

Keyword:

Novelty - main request (yes)

Inventive step - mixture of technical and non-technical features - main request (no)

Amendment to appeal case - amendment gives rise to new objections (yes)

Amendment after summons - taken into account (no)

Decisions cited:

T 0641/00, G 0001/19



Beschwerdekammern
Boards of Appeal
Chambres de recours

Boards of Appeal of the
European Patent Office
Richard-Reitzner-Allee 8
85540 Haar
GERMANY
Tel. +49 (0)89 2399-0
Fax +49 (0)89 2399-4465

Case Number: T 1898/17 - 3.4.01

D E C I S I O N
of Technical Board of Appeal 3.4.01
of 5 October 2021

Appellant: Vocollect, Inc.
(Applicant) 703 Rodi Road
Pittsburgh,
Pennsylvania 15235 (US)

Representative: Haseltine Lake Kempner LLP
Cheapside House
138 Cheapside
London EC2V 6BJ (GB)

Decision under appeal: **Decision of the Examining Division of the
European Patent Office posted on 27 March 2017
refusing European patent application No.
06734403.6 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chairman P. Fontenay
Members: T. Petelski
C. Almberg

Summary of Facts and Submissions

- I. The applicant lodged an appeal against the Examining Division's decision to refuse the European patent application 06 734 403.
- II. In its decision, the Examining Division found that the independent claims of the then main request and auxiliary request lacked novelty (Article 54 EPC) in view of document

D3: EP 1 377 000 A1.
- III. The appellant requests that the appealed decision be set aside and that a patent be granted on the basis of the main request, filed for the first time with the statement of grounds of appeal and resubmitted on 3 August 2021, or on the basis of a first or second auxiliary request, filed on 3 August 2021.
- IV. The filings of the first and second auxiliary requests of 3 August 2021 were a response to the Board's summons to oral proceedings and the communication annexed to it. In this communication, the Board set out its preliminary opinion that the claims of the main request failed for lack of an inventive step in view of D3.
- V. Claim 1 of the main request reads:

A method for recognizing speech, the method comprising the steps of:

analyzing speech input to generate a hypothesis and a confidence factor associated with the hypothesis;

comparing said confidence factor to an acceptance threshold for accepting the hypothesis; and

comparing the hypothesis to an expected response, and:

if the comparison is not favorable, then not adjusting the acceptance threshold prior to comparing the confidence factor to the thereto,

if the comparison is favorable, adjusting the acceptance threshold, prior to comparing the confidence factor thereto, in order to make acceptance of the hypothesis more likely.

VI. Claim 1 of the first auxiliary request reads:

A method for recognizing speech, the method comprising the steps of:

analyzing speech input from a user to generate features of the input speech and to generate simultaneously

by comparing the input speech features to features of an expected response using in a match or search algorithm only a model or models associated with the expected response, a first hypothesis and a first confidence factor associated with the first hypothesis, and

by comparing the input speech features to features of additional responses, a second hypothesis and a second confidence factor associated with the second hypothesis, wherein the first hypothesis has higher priority than the second hypothesis;

comparing said first confidence factor to an acceptance threshold for accepting the first hypothesis, and:

if the comparison is not favorable then rejecting the hypothesis, and

if the comparison is favorable then accepting the hypothesis; and

comparing the first hypothesis to the expected response wherein the expected response is known beforehand, and:

if the comparison is not favorable, then not adjusting the acceptance threshold prior to comparing the first confidence factor thereto,

if the comparison is favorable, adjusting the acceptance threshold, prior to comparing the first confidence factor thereto, in order to make acceptance of the first hypothesis more likely.

VII. Claim 1 of the second auxiliary request reads:

A method for recognizing speech, the method comprising the steps of:

analyzing speech input to generate feature vectors for use in searching an acoustic model to determine a hypothesis;

modifying the acoustic model based on an expected response, wherein the expected response is known beforehand;

generating the hypothesis and a confidence factor associated with the hypothesis;

comparing the confidence factor to an acceptance threshold for accepting the hypothesis, and:

if the comparison is not favorable then rejecting the hypothesis, and

if the comparison is favorable then accepting the hypothesis; and

comparing the hypothesis to the expected response, and:

if the comparison is not favorable, then not adjusting the acceptance threshold prior to comparing the confidence factor thereto,

if the comparison is favorable, adjusting the acceptance threshold, prior to comparing the confidence factor thereto, in order to make acceptance of the hypothesis more likely.

VIII. The appellant's arguments, in so far as relevant for the decision, are set out in the Reasons, below.

Reasons for the Decision

The invention

1. The application is concerned with speech recognition. More particularly, with improving the speed and accuracy of speech recognition when one or more expected responses are likely. Typically, the received speech is analyzed by extracting acoustic features and by matching the extracted features to an acoustic speech model. Thereby, the word (or words) that was likely to be spoken is identified as a hypothesis. The

likelihood of the hypothesis, corresponding to the word actually spoken, is represented by a confidence factor that is assigned to each hypothesis. The decision, whether to accept the hypothesis or not, depends on a comparison of its confidence factor with an acceptance threshold.

2. It may happen that a hypothesis, despite being correct, is not accepted, because its respective confidence factor is too low. In that case, the speaker will have to be asked to repeat the response, or to spell it.
3. It is the resulting, unnecessary loss of time, which the invention aims to prevent (cf. paragraphs [0006], [0007] and [0021] of the original application).
4. The invention is aimed at situations in which a certain speech content is expected. As an exemplary situation the application mentions inventory management (see paragraphs [0004], [0005] and [0022]). Under the command and control of a central computer a worker may perform manual tasks whilst exchanging vocal information with the computer through a headset. One such task can be the picking of items from a warehouse. Here, the worker may have to confirm location and number of the picked items. The uttered check-digit may, therefore, be expected.
5. The original application describes several different techniques that use the knowledge of an expected response in order to facilitate its recognition. Only one of those techniques, described with reference to Figure 2, was prosecuted during examination and is claimed in the main request. In this technique, the hypothesis (for example one of the numbers "one" to "six"; cf. paragraph [0047]) is compared to the

expected response (for example "one"). If the comparison is favorable (i.e. the hypothesis is "one"), the acceptance threshold is adjusted such that the acceptance of the hypothesis is rendered more likely. Thereby, it can often be avoided to prompt the user to repeat or spell the correct response.

Disclosure of D3

6. D3 lies in the field of speech recognition in automated directory services. Such a directory service may be foreseen to provide information of a specific person, like its phone number, name or address, to a caller (cf. paragraphs [0002] and [0020] of D3). An automated dialogue manager guides the caller through the enquiry by posing questions to which the caller has to reply. The caller's responses are analyzed using a speech recognizer. The latter matches the responses to the entries of a lexicon 70, which contains a list of possible answers. In the embodiment described in paragraphs [0033] and [0034], the dialogue manager asks for a town name and matches the response only to that part of the lexicon 70 that comprises town names.
7. A confidence factor (here called "confidence level" or "metric" 71) is assigned to each hypothesis that results from the matching. The confidence factor is compared to an acceptance threshold (here the "threshold T") and, depending on the result of the comparison, the hypothesis is either accepted or not (cf. paragraph [0034]).
8. D3 describes different ways of adapting the threshold to the situation: The threshold can be determined during initialization, as is described in relation to

Figure 3 (cf. paragraphs [0038] to [0043]); it can be adapted and even continually tuned during the dialogue (cf. paragraphs [0037], [0044] and [0045]); or different thresholds can be assigned to different entries of the lexicon (cf. paragraph [0046]).

Main Request - Novelty in view of D3

9. The appellant identifies the following differences between the subject-matter of claim 1 and D3:
 - (a) The invention allowed a greater variety of responses, including a variety of non-expected responses. In contrast, D3 could only recognize a limited number of expected responses, for example town names. The restricted list of vocabulary mentioned in the description of the present application in paragraphs [0047] and [0062] was not part of the invention as claimed.
 - (b) The invention defined a two-step method. A hypothesis was generated in the first step, and in the second step, the hypothesis was compared to the expected response. In contrast, the speech in D3 was directly matched to a list of expected responses in a one-step method, and no later comparison step was necessary.
 - (c) The invention defined an adjustment of the acceptance threshold depending on the above mentioned comparison. The relevant embodiment in D3, described in paragraph [0048], used a table with a different threshold for each town. The threshold for the speaker's home town was assigned after acquiring the knowledge of the home-town.

Hence, no adjustment took place, least of all as a consequence of a comparison.

10. As to (a), this alleged difference is nothing more than a difference in the number of entries in a list of possible words used for the speech recognition. In some embodiments of the invention, the speech is (implicitly) matched to a finite number of entries of a lexicon, even if that might be a lexicon of the complete vocabulary of a language in certain embodiments. In other embodiments of the invention (cf. paragraph [0047] of the description as filed), the list can be even shorter than in D3, for example comprising only the numbers 0 to 9. In D3, the speech is matched to a lexicon containing a list of names (cf. paragraph [0028]), or, in the example mentioned above, to a part of the lexicon that contains town names (cf. paragraph [0034]). Hence, the alleged difference is reflected only in some, but not all, embodiments of the invention. The claims neither restrict nor limit the matching to lists of a particular length or content.
11. Hence, no difference to D3, in the sense of a higher variety of responses, can be recognized. It follows that D3 comprises the feature "analysing speech input ...".
12. The alleged differences (b) and (c), and the respective disagreement of the appellant with the examining division, is caused by a different interpretation of one particular embodiment of D3, which is shortly mentioned in paragraph [0048].
13. In this embodiment, the threshold can depend on an *a priori* knowledge about the speaker, gained, for example, from previous dialogue elements. The *a priori*

knowledge can be the location (home-town) of the user. For example, a user located in the city of Lausanne is more likely to request an address in the same city of Lausanne. In that case, the home-town of the user (Lausanne) is an expected response. Hence, the confidence level (or confidence factor) associated to the hypothesis "Lausanne" will be matched to a favorable threshold, different from the less favorable threshold used for other hypotheses (for example the city of Lausen).

14. The Examining Division, in its decision, understands this embodiment as follows. The speaker named the town for which a more detailed address was requested. The speech recognition determined a hypothesis (for example "Lausen" or "Lausanne") and assigned it a respective confidence level. It was implicit that the hypothesis would have to be compared to the *a priori* knowledge on the speaker's home town, in other words, the expected response ("Lausanne"). It also followed implicitly that the one single threshold would have to be adjusted in case of a favorable outcome of the comparison. As a consequence, the Examining Division concluded that the comparison and adjustment steps were disclosed by D3.

15. The appellant interprets this embodiment differently. It followed from the formulation "the threshold depends on the location of the user ..." in paragraph [0048] that there were different thresholds for different towns, similarly to the embodiments described in paragraph [0046]. These thresholds were assigned, for all towns, after the acquisition of the *a priori* knowledge on the speaker's home town ("Lausanne"). This had the effect that the threshold assigned to the city of Lausanne was different from the threshold assigned to other towns. Hence, in contrast to claim 1, there

was no adjustment of a threshold, because the threshold was simply assigned for the purpose of the call, once and for all. As a second difference, no comparison of the hypothesis to an expected response ("Lausanne") was performed. Instead, the confidence factor was directly compared to the threshold value assigned to the respective town.

16. Neither of the two interpretations follows unambiguously from D3. It is simply not clear from paragraph [0048] if the embodiment refers
 - to one single, adaptable threshold (which is the finding of the Examining Division), or
 - to a table that assigns a threshold to each town (which is the opinion of the appellant), or
 - to two threshold values, one for the home-town of the speaker and one for all other towns (which would be another possible interpretation).

17. Hence, there is no explicit or implicit disclosure of the step "comparing the hypothesis to an expected response". Consequently, there is also no disclosure of a threshold adjustment that depends on the outcome of said comparison.

18. As a consequence, the subject matter of claim 1 is novel in view of D3 (Article 54 EPC). The same holds for claim 17.

Main request - Inventive step in view of D3

19. In the previous paragraph it was established that the subject-matter of claim 1 differed from D3 in the step of comparing the hypothesis to an expected response, and in the subsequent step of adjusting the acceptance

threshold in case of a favorable outcome. Comparing two parameters and, depending on the result, adjusting a third parameter is a purely mathematical operation. Therefore, the steps are non-technical by themselves. In other words, the distinguishing features are non-technical.

20. When applying the problem-solution approach, it is established case law that the problem may be formulated including non-technical (here: mathematical) features ("Comvik"-approach as set out in T 641/00 and confirmed, most recently, in G 1/19; see also Case Law, 9th Edition, I.D.9.1.3). In the current case, it will need to be established, whether said distinguishing, non-technical features interact with the technical features of the claim in so far as to contribute to a technical effect, thus contributing to the solution of a technical problem.

21. D3 achieves the same overall goal as the invention: The *a priori* knowledge of an expected response is used for setting the acceptance threshold such that it is more likely to accept a hypothesis that corresponds to the expected response. Thereby, unnecessary fallback actions, like a repetition of the response, can be avoided. This saves time. The value of the acceptance threshold is not dependent on the setting process. Namely, it does not depend on whether it is assigned after comparison of the hypothesis with the expected response, or whether it is assigned from the start, taking due account of the expected response. Hence, no effect can be derived from the value of the acceptance threshold, as such. The effect of the distinguishing steps merely lies in the way, in which the expected response is used in order to compare the confidence factor of the hypothesis with the proper acceptance

threshold. This effect does not serve any technical purpose and does, therefore, not have a technical character.

22. The objective problem can be seen as finding an implementation of how to set the proper acceptance threshold for comparison with the confidence factor of the generated hypothesis, considering the expected response.
23. The solution to this problem involves only non-technical considerations in the form of the selection of certain mathematical operations. The distinguishing, non-technical features, therefore, do not contribute to the technical character of the method defined in claim 1.
24. The same conclusion applies to the system of independent claim 17. The system requires the presence of corresponding calculating means for comparing the hypothesis to the expected response and for adjusting the threshold.
25. The system of D3 also comprises calculating means. Hence, the contribution of the claimed system over D3 is, again, limited to the mathematical operations being carried out. Since there is no contribution over D3 of a technical nature in the claimed invention, an inventive step cannot be recognised.
26. In the opinion of the appellant, the distinguishing steps identified above did have technical effects. The comparison and threshold adjustment enabled a greater variety of responses to be recognized, because the speech was not directly matched to a limited list of expected responses, as it was the case in D3. Further,

storage space could be saved, because the invention required the saving of only one threshold value, and not a separate value for every possible hypothesis. In addition, processing time and resources could be saved, because no threshold adjustment was necessary at all if the hypothesis did not match the expected response. The appellant added that the objective technical problem was to provide an alternative way of speech recognition. Speech recognition was commonly accepted to be of technical nature.

27. The arguments are not persuasive. As noted further above, the claim does not imply the variety of speech to be detected. The claim encompasses restricted speech recognition, limited for example to numbers one to six as envisaged in paragraph 47 of the application.
28. As to the storage capacity, independently of the fact that potential saving would be counterbalanced by the additional memory space required for the program incorporating the steps of comparing and adjusting, it is observed that the alleged effect cannot be derived from the claims wording. The claims wording also encompasses acceptance thresholds being defined for each possible response. That aside, D3 does not imply the presence of a large number of threshold values. There might well be only two values, one for the home-town of the user and one for the other towns.
29. There is also no apparent saving of processing time of the response. In D3, the threshold has either been set previously, after acquiring the knowledge on the caller's home town, or it is set only if the requested town corresponds to the home town. Hence, the processing of the response does not require more resources.

30. It is true that speech recognition *per se* is typically recognized as being technical. However, it is not the technical character of the claim as a whole that is put in question, but the technical contribution of the distinguishing features to the prior art, in this case to D3. As shown above, the distinguishing features are neither technical by themselves, nor do they contribute to solve a technical problem.
31. As a consequence, claim 1 of the main request does not involve an inventive step in view of D3 (Article 56 EPC).

Auxiliary Requests - Admission

32. The first and second auxiliary requests were filed in response to the Board's preliminary opinion as notified. Hence, they constitute amendments to the appeal case the admission of which is governed by Article 13(2) RPBA 2020, under which the criteria applicable under Article 13(1) RPBA 2020 may be relied on (see EPO OJ Suppl. 2/2020, Table setting out the amendments to the RPBA and the explanatory remarks, page 60).
33. The Board does not admit these requests because they are *prima facie* not allowable and give rise to new objections. The reasons are as follows.
34. According to the appellant, the features added to claim 1 of the first auxiliary request were based on original claims 55 and 56, which correspond to the embodiment described in paragraphs [0055] and [0059] to [0061] of the original application, with reference to Figure 4.

According to this embodiment, features of the input speech are compared only to features of an expected response in order to generate a hypothesis. In the corresponding example, described in paragraphs [0062] and [0063], a spoken two-digit response from a user is compared to the expected response "three five". Hence, the hypothesis necessarily corresponds to the expected response. A later comparison of the hypothesis to the expected response would not make any sense. If the confidence factor exceeds the threshold, the hypothesis is accepted directly. If the confidence factor does not exceed the threshold, the response will be compared to other models containing the remaining 99 two-digit combinations for generating another (second) hypothesis.

35. The features added to claim 1 of the second auxiliary request were based on the embodiment described in paragraphs [0067], [0068] and [0069] of the original disclosure with reference to Figure 5. Here, the acoustic model is modified based on the expected response. This is done such that using the modified model, it will be more probable that the feature matching will lead to a hypothesis that corresponds to an expected response.

36. These embodiments are fundamentally different from the embodiment that has been claimed in the main request, which refers to Figure 2. The embodiments relating to Figures 4 and 5 achieve the result of favouring the acceptance of an expected response by influencing the speech recognition, which is mutually exclusive with the step of comparing the hypothesis with an expected response and the following step of adjusting (or not adjusting) the acceptance threshold. The references to other embodiments in paragraphs [0061] and [0068],

which the appellant has identified, merely refer to non-essential features like the reception and processing of speech.

37. In contrast to the disclosure of the original application, amended claim 1 combines the mutually exclusive step of influencing the speech recognition (by using a model only associated with the expected response in auxiliary request 1 and by using a model modified based on the expected response in auxiliary request 2) and the step of comparing the (first) hypothesis to an expected response for adjusting (or not adjusting) the acceptance threshold.

38. Hence, the auxiliary requests, *prima facie*, give rise to new objections regarding at least added subject-matter (Article 123(2) EPC). In addition, the amendments shift the scope of the claims to previously non-claimed subject-matter (influencing the initial speech recognition), which lies outside of the invention claimed so far in the course of the examination and appeal proceedings. Although claim 1 according to auxiliary requests 1 and 2 reproduces the features of claim 1 of the main request, the shift resulting from the introduction of the features of the alternative embodiments of figures 4 and 5 amounts to introducing subject-matter into the appeal proceedings that was not pursued when the application entered the European phase. Their filing is tantamount to creating fresh cases. This is not acceptable (see, in addition, Article 12(2) RPBA 2020).

39. As a consequence, the Board uses its discretion not to admit those requests into the proceedings.

Summary of conclusions

40. The main request is not allowed, and the first and second auxiliary requests are not admitted into the proceedings.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar:

The Chair:



H. Jenney

P. Fontenay

Decision electronically authenticated