Internal distribution code:

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

Datasheet for the decision of 17 December 2025

Case Number: T 1215/23 - 3.4.01

Application Number: 18874826.3

Publication Number: 3686882

IPC: G10L15/00, G10L15/065, G10L15/20

Language of the proceedings: EN

Title of invention:
METHOD FOR TRAINING FILTER MODEL AND SPEECH RECOGNITION METHOD

Applicant:
Shenzhen Yinwang Intelligent Technologies Co., Ltd.

Headword:
Speech recognition / Shenzhen Yinwang

Relevant legal provisions:
EPC Art. 84

Keyword:
Main request and first to third auxiliary requests - clarity (no)

Europäisches Patentamt
European Patent Office
Office européen des brevets

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

Case Number: T 1215/23 - 3.4.01

D E C I S I O N
of Technical Board of Appeal 3.4.01
of 17 December 2025

Appellant: (Applicant)	Shenzhen Yinwang Intelligent Technologies Co., Ltd. Room 101, Huawei Headquarters Office Building Huawei, Vanke City Community Bantian Street Longgang District Shenzhen City, Guangdong 518129 (CN)
Representative:	Gill Jennings & Every LLP The Broadgate Tower 20 Primrose Street London EC2A 2ES (GB)
Decision under appeal:	**Decision of the Examining Division of the European Patent Office posted on 21 February 2023 refusing European patent application No. 18874826.3 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chair	P. Scriven
Members:	T. Petelski
	L. Bühler

Summary of Facts and Submissions

I. The appeal is of the Examining Division's decision to refuse the application for lack of clarity of the claims of the main request and all three auxiliary requests.

II. The appellant requested that the decision be set aside and a patent granted on the basis of the main and three auxiliary requests underlying the contested decision and re-filed on appeal. Oral proceedings were requested, should the main request not be allowed.

III. In an annex to a summons to oral proceedings, the Board informed the appellant of its preliminary opinion, according to which the Examining Division was right in finding that claim 1 of the main request and all three auxiliary requests lacked clarity.

IV. In a further submission, and during oral proceedings, the appellant defended their position, maintaining their initial requests.

V. Claim 1 of the main request reads:

> *A filtering model training method, wherein the filtering model is used when performing speech recognition processing, wherein the method comprises:*

determining N original syllables, wherein the N original syllables are syllables comprised in an actual pronunciation of a first corpus, and N is an integer greater than or equal to 1;

determining N recognized syllables, wherein the N recognized syllables are syllables of a recognition result obtained after first speech recognition processing is performed on sound signals of the first corpus, the first speech recognition processing comprises filtering processing based on the filtering model and recognition processing based on a speech recognition engine, and the N recognized syllables are in a one-to-one correspondence with the N original syllables;

determining N syllable distances based on the N original syllables and the N recognized syllables, wherein the N syllable distances are in a one-to-one correspondence with N syllable pairs, the N original syllables and the N recognized syllables form the N syllable pairs, each syllable pair comprises an original syllable and a recognized syllable that correspond to each other, and each syllable distance is used to indicate a similarity between an original syllable and a recognized syllable that are comprised in a corresponding syllable pair; and

training the filtering model based on the N syllable distances;

wherein the training the filtering model based on the N syllable distances comprises:

determining a scoring model based on the sound signals of the first corpus and the N syllable distances, wherein an output of the scoring model represents a distance between a recognized syllable and an original syllable;

determining K syllable distances based on the scoring model and sound signals of a third corpus, wherein an actual pronunciation of the third corpus comprises K original syllables, a recognition result obtained after the first speech recognition processing is performed on the sound signals of the third corpus comprises K recognized syllables, the K recognized syllables are in a one to-one correspondence with the K original syllables, the K syllable distances are in a one-to-one correspondence with K syllable pairs, the K original syllables and the K recognized syllables form the K syllable pairs, each syllable pair comprises an original syllable and a recognized syllable that correspond to each other, each syllable distance is used to indicate a similarity between an original syllable and a recognized syllable comprised in a corresponding syllable pair, and K is an integer greater than or equal to 1; and

training the filtering model based on the N syllable distances and the K syllable distances.

VI. Claim 1 of the first auxiliary requests adds, to the end of claim 1 of the main request, the following:

[... the K syllable distances];

wherein the training the filtering model based on the N syllable distances comprises:

determining environment information of an environment in which the filtering model is used; and

training the filtering model based on the N syllable distances and the environment information.

VII. Claim 1 of the second auxiliary requests further adds the following:

[... environment information];

wherein when the filtering model is configured in a vehicle, the environment information comprises at least one of the following information:

vehicle speed information, information about whether a vehicle window is open or closed, or air volume information of an air conditioner.

VIII. Claim 1 of the third auxiliary requests even further adds the following:

[... an air conditioner];

wherein each syllable comprises at least one phoneme; and

the determining N syllable distances based on the N original syllables and the N recognized syllables comprises:

obtaining first mapping relationship information, wherein the first mapping relationship information is used to indicate a phoneme distance between a plurality of phonemes, and a phoneme distance between any two phonemes is used to indicate a similarity between any two phonemes; and

determining the N syllable distances based on the first mapping relationship information.

Reasons for the Decision

Content of the application

1. As far as can be gathered from the description, the invention seeks to improve speech recognition in certain environments by using a filtering model to pre-filter the sound signal prior to feeding it to the speech recognition engine. The invention is, in particular, concerned with training the filtering model based on the phonetic distances between syllables

recognized by the speech recognition engine in a filtered sound signal from a known corpus of spoken or played text, and the correct syllables of that corpus. The trained filtering model is adapted both to the environment and to the speech recognition engine.

Main request - clarity

2. According to the Examining Division, the description used the term "syllable" in a sense that went far beyond its well established meaning in linguistics, covering any unit of speech, from a single phoneme to multiple words. This was a non-standard use of the term, which rendered the claims, which used the term in various places, unclear.

3. In a communication under Rule 15(1) RPBA, the Board laid out its preliminary opinion, according to which the skilled person would understand the term "syllable" according to its well-established meaning, as defined in paragraph [0069] of the application, despite the confusion in the description between the terms "phoneme" and "syllable". Hence, the inconsistency in the use of these terms in the description was not one that meant that "syllable" was unclear in the claims.

4. In its reply, the appellant explained that the Board's understanding was not correct, as what is referred to as syllables differs between languages.

5. Hence, the Board is led to change its opinion, meaning the Examining Division might well have been correct in finding that the inconsistent use of the term "syllable" in the description rendered claim 1 unclear.

However, in view of the other objections, this issue can be left unsolved.

6. The Examining Division also found that the term "filtering model", which appeared repeatedly in the claims, did not have a well-defined meaning, nor was it sufficiently defined in the claims. In addition, the claims lacked a definition of the role of the "filtering model" in the speech recognition process, and of the way in which it was trained. This also rendered the claims unclear.

7. The appellant argued that the core of the invention was to use a filtering model before sending the recorded speech signals to the speech recognition engine, and to train the filtering model based on syllable distances such that the output of the speech recognition engine was closer to the known text of the corpus. The details of the filtering model, and of the scoring model that was also used in the training, were not what the invention was about. Accordingly, their realization was left to the technical knowledge of the skilled person.

8. The appellant added that the application provided sufficient information on both the filtering and the scoring model. For example, Figure 5 and paragraph [0196] of the application illustrated the nature of the models and their interrelation. The skilled person would understand that the N syllable distances represented an output of the filtering model. The N syllable distances were also used to determine the scoring model. Therefore, the K syllable distances, having been determined based on the scoring model, represented a measure of success of the filtering model. The claim did not restrict the training of the filtering model to one based only on the N syllable

distances. Rather, it could also be based on other things, like the K syllable distances.

9. The appellant concluded that the skilled person would understand how to realize a suitable filtering model and scoring model, and there was no need for a more detailed definition in claim 1.

10. The Board is not persuaded by these arguments. Instead, it shares the Examining Division's view that the nature of the filtering model and its training are unclear in claim 1. This will be explained in the following.

11. The skilled person understands that the fact that the filtering model is trained implies that it is subject to machine learning; hence, that training data are fed to the model, and that the output, after being further processed by the speech recognition engine, is compared to the correct results (assumed to be known), a measure of success is determined, and certain variables of the model are repeatedly adjusted in order to maximize the success (minimize the error).

12. Claim 1 implies that the trained model will be used to filter an unknown sound signal before actual speech recognition is performed on the filtered signal. The training is expected to alter the model in such a way that its trained version leads to improved speech recognition.

13. As the appellant rightly noted, a model subjected to machine learning does not necessarily require a definition of its nature, or of the exact training process, if common general knowledge allows the skilled person to understand which model to use and how to train it. For example, the skilled person might know

what kinds of model respond to training with environmental noise data.

14. However, the present case is different in that the training is based on syllable distance information; for example, the information that syllable "ka" might have a large syllable distance from its recognized counterpart "ke". Here, without further explanation, it is not clear what kind of filtering model could be trained based on such information, and how such training could be performed in order to modify the model to improve the subsequent speech recognition.

15. The steps of the "filtering model training method" defined in claim 1 contradict the assumptions of the skilled person regarding the training, and further obscure the nature of the filtering model and its training.

- In step 1, N original (known) syllables of a first corpus (the training data) are determined.

- In step 2, N recognized syllables corresponding to the N original syllables are determined as a result, firstly, of filtering the sound signals of the first corpus using the filtering model, and, secondly, of processing the signals using a speech recognition engine. This step uses the (untrained, or partially trained) filtering model for determining the sample data (recognized syllables) that are to be compared to the correct results (original syllables) determined in step 1.

- In step 3, the N syllable distances between the N syllable pairs of the syllables determined in steps 1 and 2 are determined. This step implicitly

involves a comparison of the sample data (recognized syllables) with the correct results (original syllables).

- Step 4 performs "training the filtering model based on the N syllable distances." One might have expected that the training would encompass repeatedly adjusting variables of the filtering model and performing speech recognition on the filtered sound signals from various training corpora, until the syllable differences of the recognized speech were within a desired range. Instead, step 4 defines the training by the following three sub-steps:

  - In sub-step 4a, a "scoring model" is determined based on the sound signals of the first corpus (the training data) and the corresponding syllable distances of step 3. The output of the scoring model represents one of the syllable distances on which the scoring model is based.

  - In sub-step 4b, a number, K, of syllable distances of syllables of a different, third corpus are determined. For determining the K recognized syllables, only "speech recognition processing" is performed on the sound signals of the third corpus, but no filtering. Somehow, the determination of the K syllable distances is not only based on the sound signals, but also on the scoring model determined in step 4a.

  - In sub-step 4c, the filtering model is trained based on the N syllable distances of the first corpus and the K syllable distances of the third corpus.

16. There are clarity problems relating to sub-steps 4a, 4b, and 4c.

17. In step 4a, it is not clear, in how far the sound signals of the first corpus (the training data) can be involved in determining the scoring model in addition to the N syllable distances, considering that the output of the scoring model represents a distance between a (single) syllable pair. Furthermore, the nature and purpose of the scoring model is not clear, because the distance between a single syllable pair is hardly suitable to serve as a measure of success of the speech recognition of the (entire) first corpus. Indeed, contrary to what the term "scoring model" suggests, the claim does not define the output of the scoring model as a measure of success ("score"), which is used in an (iterative) training of the filtering model. Rather, it appears that the N syllable distances serve as the measure of success, on which the training is based.

18. According to feature 4c, the training of the filtering model is also based on K syllable distances of a third corpus. As the K syllable distances can be determined without using the filtering model at all, it is not clear, how the filtering model can be trained on that basis.

19. Sub-step 4b defines that the determination of the K syllable distances related to the third corpus is based on the scoring model. Apart from confirming that the output of the scoring model is not used as a "score", it is unclear, from this definition, how the output of the scoring model, which represents a syllable distance derived from the first corpus, could be used in determining the K syllable distances of the third

corpus. This is all the more the case, considering that the third corpus could contain entirely different syllables than the first corpus. This further obscures the nature of the scoring model and its role in the training of the filtering model.

20. The unclear training steps, including the unclear nature and role of the scoring model, are not suited to defining the nature of the filtering model and its training, which appears to be essential to properly define the invention.

21. Hence, the Examining Division was right in finding that the subject-matter of claim 1 is not clear (Article 84 EPC), which is why the main request is not allowable.

Auxiliary requests - clarity

22. Claim 1 of the first auxiliary request additionally includes the features of claim 8 of the main request, which define the determination of environment information, which is also used in training the filtering model.

23. Claim 1 of the second auxiliary requests further adds the features of claim 9 of the main request, which define particular environmental information under the condition that the filtering model is configured in a vehicle.

24. Claim 1 of the third auxiliary requests further adds the features of claim 4 of the main request, which define that each syllable comprises at least one phoneme, and that the determination of the N syllable distances happens via a mapping relationship

information indicating a phoneme distance between a plurality of phonemes.

25. With regard to Auxiliary Requests 1 and 2, the appellant argued that the use of environmental information in the training of the filtering model clarified that the filtering model was used to account for the environment in which the speech recognition was to be performed. This was even clearer for the case in which the filtering model was trained for the environment of a vehicle.

26. However, even if the skilled person knew how to train a filtering model based on environmental information alone, using environmental information in addition to syllable distances has no effect on the clarity issues resulting from the latter. In particular, the definition of the environmental information does not clarify the nature of the filtering model or how it is trained based on syllable distances. Nor does it clarify the nature of the scoring model or its role in training the filtering model, nor the training of the filtering model based on the K syllable distances of the third corpus. Therefore, claim 1 of each of the second and third auxiliary request remains unclear.

27. The further amendment to Auxiliary Request 3 was introduced in response to a clarity objection raised, by the Examining Division, with regard to the term "syllable distance". However, since the Board's objections are not based on the clarity of this term, and since the amendment has no bearing on the clarity of the "filtering model", the amendment does not change the situation with regard to the higher-ranking requests.

28. Hence, the Examining Division was right in finding that the subject-matter of claim 1 of each of the auxiliary requests suffers from the same clarity problems related to the "filtering model" as the main request (Article 84 EPC), which is why these requests are also not allowable.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar: The Chair:

D. Meyfarth P. Scriven

Decision electronically authenticated