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 19 November 2007

Case Number: T 1002/04 - 3.4.01

Application Number: 99307658.7

Publication Number: 0992980

IPC: G10L 15/26

Language of the proceedings: EN

Title of invention:
Web-based platform for interactive voice response (IVR)

Applicant:
LUCENT TECHNOLOGIES INC.

Opponent:
-

Headword:
-

Relevant legal provisions:
EPC Art. 52(1), 56
RPBA Art. 11(3)

Keyword:
"Inventive step - no"

Decisions cited:
-

Catchword:
-

Europäisches Patentamt | European Patent Office | Office européen des brevets

Beschwerdekammern | Boards of Appeal | Chambres de recours

Case Number: T 1002/04 - 3.4.01

D E C I S I O N
of the Technical Board of Appeal 3.4.01
of 19 November 2007

Appellant: LUCENT TECHNOLOGIES INC.
600 Mountain Avenue
Murray Hill, NJ 07974-0636 (US)

Representative: Sarup, David Alexander
Alcatel-Lucent Telecom Limited
Unit 18, Core 3
Workzone
Innova Business Park
Electric Avenue
Enfield, EN3 7XU (GB)

Decision under appeal: **Decision of the Examining Division of the European Patent Office posted 23 April 2004 refusing European application No. 99307658.7 pursuant to Article 97(1) EPC.**

Composition of the Board:

Chairman: B. Schachenmann
Members: F. Neumann
G. Assi

Summary of Facts and Submissions

I. The appellant (applicant) lodged an appeal, received on 17 June 2004, against the decision of the examining division, dispatched on 23 April 2004, refusing European patent application No. 99 307 658.7 (publication number 0 992 980). The fee for the appeal was paid on 14 June 2004. The statement setting out the grounds of appeal was filed on 30 July 2004.

II. In the contested decision, the examining division held that the subject-matter of claim 1 on file at that time lacked an inventive step (Articles 52(1), 56 EPC).

III. With the notice of appeal, the appellant requested that the decision under appeal be set aside and a patent granted. With the statement of grounds of appeal, the appellant filed an auxiliary request that a patent be granted on the basis of a set of claims marked "AUXILIARY SET".

IV. Reference is made to the following documents:

D1: KAZUHIRO KONDO ET AL: "A WWW BROWSER USING SPEECH RECOGNITION AND ITS EVALUATION", SYSTEMS & COMPUTERS IN JAPAN, US, SCRIPTA TECHNICAL JOURNALS. NEW YORK, vol. 29, No. 10, 1 September 1998 (1998-09-01). pages 57-66, XP000786722 ISSN: 0882-1666;

D3: EP-A-0 848 373

D4: WO 98/35491.

V. In an annex to a summons to oral proceedings, dispatched on 23 July 2007, the Board set out its

preliminary opinion that claim 1 of the main request lacked an inventive step in view of a combination of D1 and D3.

VI. In a reply dated 3 September 2007, the appellant informed the Board that he would not attend the oral proceedings and requested that the proceedings be continued in writing. It was requested that a patent be granted on the basis of claims 1-15 filed with the letter of 3 September 2007; the auxiliary request was withdrawn.

VII. Oral proceedings took place in the absence of the appellant on 18 October 2007.

VIII. The wording of independent claim 1 reads as follows:

"A method for implementing an interactive voice response application over a network (104, 109), the method comprising the steps of:

generating in a speech synthesizer (116) speech output characterizing at least a portion of a web page retrieved over the network;

processing information in the web page in a grammar generator (120) to produce at least a portion of a grammar;

utilizing the grammar to recognize speech input in a speech recognizer (122) having an input coupled to an output of the grammar generator; and

utilizing the grammar in the speech synthesizer to create phoneme information;

wherein the phoneme information created by the speech synthesizer utilizing the grammar is provided to the speech recognizer, and said phoneme information is

used in both the recognizing of said speech input in the speech recognizer and the generating of said speech output in the speech synthesizer."

IX. The appellant essentially relied on the following submissions:

In a method for implementing a web-based interactive voice response (IVR) system in accordance with claim 1, there is an explicit connection between the recognition of speech input and the generation of speech output characterising a web page: this explicit connection is based on the utilisation of phoneme information derived from a given grammar. In particular, claim 1 recites that phoneme information is created by a speech synthesiser utilising a grammar generated by processing information in the web page, and that the phoneme information is provided to a speech recogniser that also utilises that same grammar. The phoneme information is used in both recognising speech input in the speech recogniser and generating speech output in the speech synthesiser.

The addition of a speech generation capability to the system of D1 would not necessarily include an explicit connection between a grammar generator and both a speech recogniser and a speech synthesiser. It is possible that speech recognition and speech generation can be provided independently of each other in an IVR system without the claimed connection as is evidenced by the disclosure of D4. D4 shows a system having both a speech recogniser and a speech synthesiser in which there is no connection between the grammar generating element and the speech synthesiser.

The subject matter of claim 1 is therefore not obvious in view of a combination of D1 and D3.

Reasons for the Decision

1. The appeal is admissible.

2. D1, which is considered to represent the closest prior art, discloses a system in which navigation of visually displayed web pages is performed by spoken commands. In particular, this system enables the user to navigate through a web page by speaking the anchor text associated with embedded hyperlinks.

2.1 Using the terminology of claim 1, D1 discloses:
a method for implementing an interactive voice response application over a network (see page 57, LH column, lines 1-5; page 57, RH column, lines 11-14; Fig. 2) the method comprising the steps of:
processing, in a grammar generator (see Fig. 2: the "Grammar construction" unit), information in a web page retrieved over the network to produce at least a portion of a grammar (page 60, section 3.1, lines 4-12; Page 61, LH column, lines 27-28);
utilizing the grammar to recognize speech input in a speech recogniser (page 62, RH column, lines 4-7; Fig. 2) having an input coupled to an output of the grammar generator (see Fig. 2: the input of the "Speech Recognition" unit is coupled - albeit indirectly via the "Text-to-phone" conversion" unit - to the output of the "Grammar construction" unit); and

utilizing the grammar to create phoneme information (page 61, RH column, lines 27-29);
wherein the phoneme information created from the grammar is provided to the speech recogniser (Fig. 2; page 62, RH column, lines 4-6) and said phoneme information is used in the recognising of speech input in the speech recogniser (Fig. 2; page 62, RH column, lines 6-14).

In this feature analysis of claim 1, it is noted that the sentence-level grammars produced in the "Grammar construction" unit of D1 from the anchor texts have been equated with the "at least a portion of a grammar" in claim 1. Using this interpretation, it can be seen that the grammar of D1 (a portion of which is represented by the sentence-level grammars) is used to recognise speech input in a speech recogniser which is (indirectly) coupled to the output of the "Grammar construction" unit (see Fig. 2; page 62, RH column, lines 4-7) and that this grammar is also used to create phoneme information (as is apparent from Fig. 2).

Moreover, the term "phoneme information" in claim 1 is so broad that the "phonetic-level grammars" produced in the text-to-phone converter of D1 (see Fig. 2 of D1) are considered to fall under this expression. Page 61, RH column, lines 27 to 29 of D1 makes clear that these "phonetic-level grammars" define the phonetic strings representing the pronunciation of the anchor texts, i.e. the phonetic transcriptions of the anchor texts. As can be seen from Fig. 2 of D1, this "phoneme information" is provided to the speech recogniser and is used in the recognising of speech input in the speech recogniser.

2.2 Claim 1 of the main request is distinguished from the disclosure of D1 in that:

(i) speech output characterizing at least a portion of the web page is generated in a speech synthesiser;
(ii) the phoneme information which is created utilising the grammar, is created in the speech synthesiser; and
(iii) the phoneme information created by the speech synthesiser utilising the grammar is used not only in the recognising of the speech input in the speech recogniser (as in D1) but also in the generating of the speech output in the speech synthesiser.

2.3 Two independent technical effects are achieved by these distinguishing features. The technical effect of feature (i) is that a verbal rendering of a web page is produced. The technical effect of features (ii) and (iii) is that the recognition and synthesis branches of the system share the phoneme information generator of the speech synthesiser.

2.4 Thus the objective technical problem to be solved by the subject matter of claim 1 may be seen as the modification of the method and architecture of D1 to permit output of information from a web page in a non-visual manner whilst minimising the number of processing units employed.

2.5 D3 recognises that conventional web browsers are difficult to use in an environment where a computer monitor and/or a keyboard is/are not available. The solution proposed in D3 to this problem (see col. 1,

lines 21-27) is to provide a system in which audio playback is substituted for visual feedback and the virtual manipulation of graphical user interface (GUI) elements can be replaced by verbal input of standard browsing commands (e.g. "follow" or "scan forward" - see col. 10, lines 20-24).

2.6.1 Since D3 not only recognises the problem that visual rendering of a web-page is not always appropriate, but also provides a solution thereto, the skilled person would consider incorporating the teaching of D3 in the system of D1 and would thereby replace the visual display of D1 by the verbal rendering system of D3. This combination would result in a system which verbally renders the contents of a web page to a user and enables the user to navigate through the web pages by speaking the anchor texts, bookmark text and browser commands.

2.6.2 The verbal rendering system of D3 uses a text-to-speech (TTS) engine (see col. 4, lines 23-29, 48-52; col. 5, lines 30-34) which generates speech output characterizing at least a portion of a web page retrieved over the network (see col. 2, lines 45-5; col. 4, lines 23-29). This TTS engine of D3 necessarily includes a text-to-phoneme converter which assigns phonemic transcriptions to the textual words appearing in the web page. The skilled person, modifying the system of D1 to include the verbal rendering system of D3 would immediately recognise that both the speech recognition branch and the speech synthesis branch of the combined system are provided with their own text-to-phoneme converter: straightforward design considerations would prompt the skilled person to avoid

a duplication of processing components and thus to provide just one single text-to-phoneme converter which can be accessed by both branches. This single converter could either be provided as an independent unit or the converter from one of the branches could be made accessible to the other branch. Neither of these options are considered to be inventive as they only represent standard design alternatives. Therefore the common use of the text-to-phoneme converter of the speech synthesiser by both the speech generation branch and the speech recognition branch cannot be considered to contribute to an inventive step.

Hence, when replacing the visual display of D1 by the verbal rendering system of D3, the skilled person would dispense with the "Text-to-phone conversion" unit of D1 and would use the text-to-phoneme converter of the TTS synthesiser to perform the phonetic transcription of the text contained in the "Grammar construction" and the "Bookmarks, Browser commands" units of D1.

The Board is of the opinion that the fact that D4 discloses a system in which the speech synthesis branch and the speech recognition branch are entirely separate would not discourage the skilled person from using a shared phoneme generator as set out above. In D4, the output of the recogniser network generator 18 is made up of a vocabulary and a grammar. The vocabulary corresponds to a set of models or templates, one for each word to be recognised, and the grammar corresponds to a set of stored parameters which define the permissible word sequences (see D4, page 6, lines 1-17). Phonetic transcriptions of the text of the web page are not created in the recognition branch of

D4. There is therefore no reason to provide a link between the recogniser network generator 18 and the TTS synthesiser 15 of D4. The skilled person would not see the absence of a connection between the two branches of D4 as a disincentive to share a common phoneme generator in a configuration in which phonemes are created in each branch.

2.6.3 Claim 1 further defines that the phoneme information which is created from the grammar is used in both the recognising of speech input and the generating of speech output. The creation of the phonetic-level grammars in D1 involves the generation of the phonetic transcriptions of the anchor texts. The verbal rendering process of D3 involves the generation of a phonetic transcription of the complete text of the web-page, which text of course also contains the anchor texts. Irrespective of whether the phonetic transcription is made for speech generation or for speech recognition purposes, a phonetic transcription of the anchor texts will be required in both the speech recognition branch and in the speech generation branch. Consequently the "phoneme information" derived from the grammar (which has been equated with the phonetic transcriptions of the anchor texts in D1) will be used both in the recognising of speech input in the speech recogniser and in the generating of speech output in the speech synthesiser.

2.7 Thus, a combination of the teachings of D1 and D3, and the obvious recognition that only a single text-to-phoneme converter would be required, would result in a method for implementing an IVR application over a network as defined in claim 1. For the reasons set out

above, the skilled person would arrive at this subject matter without the use of an inventive step (Articles 52(1), 56 EPC).

3. In accordance with Article 11(3) of the Rules of Procedure of the Boards of Appeal (RPBA), the Board shall not be obliged to delay any step in the proceedings, including its decision, by reason only of the absence at the oral proceedings of any party duly summoned who may then be treated as relying only on its written case. The summons to oral proceedings was issued for reasons of expedience. Despite the further request of the appellant to cancel the oral proceedings and to continue the procedure in writing, the Board nevertheless continued with the oral proceedings - in the absence of the appellant - such that a decision could be reached in the present case. As set out in Article 11(3) RPBA, the absence of the appellant does not oblige the Board to delay its decision, which, in the present case is based on the ground of lack of inventive step on which the appellant has had an opportunity to present its comments in writing (Article 113(1) EPC).

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar: The Chairman:

R. Schumacher B. Schachenmann