(72) Inventor: **Aldrovandi, Matteo**
   40122 Bologna (IT)

(74) Representative: **Giustini, Delio**
   Siemens Mobile Communications S.p.A.
   Postfach 22 16 34
   80506 München (DE)

(54) **Voice activity detection operating with compressed speech signal parameters**

(57)   There is provided a voice activity detector (VAD) (40) for assisting the voice quality enhancement in the uplink direction of a mobile communication system in which the voice quality enhancement means (50) are embodied in the transcoding and rate adapting unit (TRAU) (2). The VAD (40) comprises means (41, 42) for performing both a spectral analysis and an energetic analysis on a received speech signal and means (43, 45) for processing the results of said analysis and taking a decision on audio segment nature. The VAD performs spectral analysis directly on the coded signal.
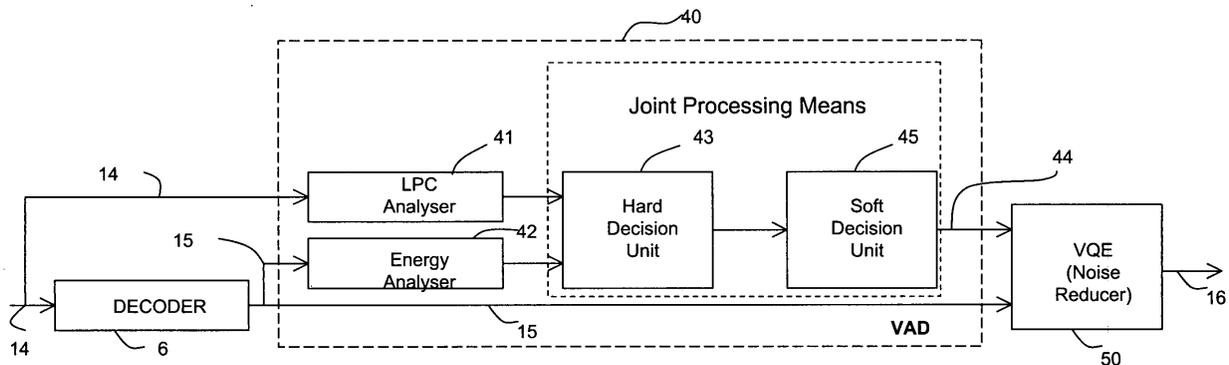
FIG. 2

EP 1 557 820 A1

Description

Field of the invention

[0001] The present invention refers to digital radio communication systems, in particular mobile communication systems, and more specifically it concerns a method of and device for voice activity detection in received speech signals in one such system.

[0002] Preferably, but not exclusively, the method and the device are intended for use in connection with voice quality enhancement.

Background of the invention

[0003] Voice activity detectors (VADs) are devices that are supplied with a signal to detect therein periods of speech and periods of silence, where only noise is present. Possibly, the VADs are also arranged to distinguish among voiced/unvoiced sounds in speech periods.

[0004] A class of VADs performs detection through an energetic analysis and a spectral analysis of the input signal, the analysis results being combined to provide the classification of an analysed speech segment. An algorithm for classifying a speech segment as voiced speech, unvoiced speech or silence based on energetic and spectral analyses is disclosed in "Application of an LPC Distance Measure to the Voiced-Unvoiced-Silence Detection Problem", by L.R. Rabiner and M.R. Sambur, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-25, No. 4, August 1977, pages 338 - 343.

[0005] In mobile communication systems, VADs are typically used at the mobile terminals, in association with the speech coder, to drive a discontinuous transmission in which coded speech signals are transmitted during active speech periods whereas in silence periods the speech transmitter is inhibited and the so-called comfort noise is transmitted. This helps in saving power.

[0006] It has now been found that it is advantageous to use a VAD also in the control part of a mobile communication system, in particular for improving the Noise Reduction (NR) feature of the so-called Voice Quality Enhancement (VQE) function. An example of VAD-assisted noise reduction for the uplink direction of communication of a digital radio communication system is disclosed in EP-A 1 017 042.

[0007] The Applicant, as well as other manufacturers, integrate the VQE function in the units (such as the Transcoding and Rate Adapting Units, or TRAU, of the GSM system) adapting the speech signals from the requirements of the radio part of the system to the requirements of the control part and, if necessary, of the fixed telephone network, and vice versa. In the Applicant's VQE, the VAD is intended to drive the noise suppression in the uplink direction of communication. For several reasons, including cost and size of the apparatus, the

Applicant wishes that the addition of a VAD-driven VQE in a TRAU does not entail changes in the hardware of the TRAU itself.

[0008] If the detection exploits both the spectrum and the energy characteristics of the received signal, the conventional approach, which is substantially as disclosed in the above-mentioned document by L.R. Rabiner and M.R. Sambur, entails that the speech signal is decoded and the spectral information (here the linear prediction coefficients LPCs) is recovered from the linear frames resulting from speech decoding. Yet, the spectral analysis of the linear signal to recover the LPCs is a heavy processing task. Taking into account that the TRAU processor generally operates in parallel on a plurality of channels, real time execution of the complete VAD algorithm on the same channels could compel to use a dedicated processor or a more powerful and hence more expensive processor than the one that would be used for the TRAU. Both solutions are in contrast with the goal of keeping the TRAU hardware unchanged.

[0009] To avoid the need for a VAD-dedicated or a more powerful processor, the spectral analysis could be dispensed with and the VAD could perform only the energetic analysis. Such a solution is disclosed in EP-A 1 017 042. The document teaches also that the energy estimation can be performed directly on the compressed signal, in order to dispense the speech decoder with the relevant processing tasks and to speed up the actual speech decoding.

[0010] Yet, by performing only the energetic analysis, only one feature of the received signal is exploited, and the detection, and hence the operation of the VAD-driven VQE, is less effective.

Object of the Invention

[0011] Thus, it is an object of the invention to provide a method and a device for voice activity detection, in particular intended to drive a voice quality enhancement integrated in a unit performing speech rate and coding adaptation in mobile communication systems, which method and device allow performing both the energetic and the spectral analysis by using the same processor as required for performing said adaptation.

Summary of the Invention

[0012] According to the invention, there is provided a method of detecting voice activity in a received speech signal in a radio communication system in which speech signals are transmitted in digitally coded form, and a signal representative of the presence or absence of voice activity is generated by submitting the received speech signals to an energetic analysis and a spectral analysis, said spectral analysis being performed directly on coded speech signals.

[0013] The invention also concerns a device for car-

rying out the method, comprising means for performing an energetic analysis and a spectral analysis on the received speech signal, in which said spectral analysis means are connected directly with a detector input where said coded speech signals are present.

[0014] In the preferred application, the voice activity detector drives a noise reduction operation, within a voice quality enhancement function performed on speech signals propagating in the uplink communication direction in a mobile communication system and embodied in units, like the so-called TRAU (Transcoding and Rate Adapting Unit), which adapt the uplink directed speech signals to the requirements of the control part of the mobile system and possibly of the fixed telephone network and adapt downlink directed speech signals to the requirements of the radio part of the mobile system.

[0015] Therefore, the invention provides also a method of voice quality enhancement in a mobile communication system, in which a voice quality enhancement including a noise reduction operation is performed at least for speech signals propagating in uplink direction, in which said noise reduction operation is driven by a signal representative of the presence or absence of voice activity generated by a method of and device for voice activity detection as defined above.

Brief description of the drawings

[0016] A preferred embodiment of the invention, given by way of non-limiting example, will now be described with reference to the accompanying drawings, in which:

-   Fig. 1 is a schematic block diagram of a TRAU embodying a VQE unit and of its connections inside the mobile communications system; and
-   Fig. 2 is a schematic block diagram of the invention.

Description of the preferred embodiment

[0017] The preferred embodiment disclosed here concerns a VAD intended to drive the noise reduction feature in a voice quality enhancement performed in the uplink direction of communication in a mobile communication system, in case the VQE function is incorporated into the units performing transcoding and/or rate adaptation in the control part of such a system.

[0018] Referring to Fig. 1, there is schematically shown a Transcoding and Rate Adapting Unit (TRAU) 1 of a mobile communication system, for instance a GSM system. The TRAU is connected to the Mobile Switching Centre (MSC) 2 and the Base Station Controller (BSC) 3 through interfaces A and Asub and embodies a VAD-driven Voice Quality Enhancement function performed in block 4 labelled VAD & VQE.

[0019] In the most general case, a VQE includes the well-known features of Acoustic Echo Cancellation, Noise Reduction and Acoustic Level Control. In the preferred application of the invention, all of said features are provided for the uplink direction of communication only and the VAD drives the Noise Reduction (NR) feature. In downlink direction, only the Acoustic Level Control is performed, which is not concerned by the present invention.

[0020] The drawing only shows the units that, in TRAU 1, are directly concerned with the transcoding function, namely a speech coder 5 and a speech decoder 6 on the Asub-interface side, and an A/μ law expander 7 and an A/μ law compander 8 on the A-interface side.

[0021] In downlink direction, the TRAU receives A-law PCM signals from MSC through a line 10, sends the expanded signals to the In_Down input of VAD&VQE block 4 through line 11. The signals outgoing from the Out_Down output of block 4 are fed to coder 5 through line 12, are coded according to the desired coding technique (full-rate, enhanced full-rate, half-rate or adaptive multi-rate) and the coded signals are then forwarded to base station controller 3 through line 13.

[0022] In uplink direction, the coded signals arriving from base station controller 3 through line 14 are fed to both decoder 6 and VAD&VQE block 4. The decoded signals are fed to the In_Down input of VAD&VQE block 4 through line 15. The decoded signals having undergone voice quality enhancement are fed from Out_Up output of VAD&VQE to A/μ law compander 8 through line 16 and hence to MSC through line 17.

[0023] It is not necessary to provide here details on the organisation of the coded speech signals in a mobile communication system, which depends on the kind of system and on the chosen coding rate. On the other hand, for any given system and rate, such organisation is well known to the skilled in the art and can be found in the relevant standards. It is sufficient here to recall that the coded speech signals include spectral information, such as the LPCs or a representation thereof.

[0024] In Fig. 2 block VAD&VQE is decomposed into its constituent blocks, namely VAD 40 and VQE 50. VAD 40 has been schematised by a spectral analyser 41 determining the LPC coefficients, an energy analyser 42 and Joint Processing Means including including an Hard Decision Unit 43 and a Soft Decision Unit 45. Said Joint Processing Means being adapted to combine the results of the two analyses and emitting on line 44 a signal indicating the nature of the received speech frame (the so-called VAD flag), which is an input to the Noise Reduction feature of VQE 50.

[0025] According to the invention, LPC analyser 41 is directly fed with the coded speech signal frames present on line 14, whereas energy estimator 42 is fed with the decoded signal outgoing from decoder 6 through line 15. The LPC analysis of course depends on the manner in which the LPCs are represented in the coded signal. The energy evaluation and the decision may be performed according to any technique known in the art, for instance as disclosed in the above-mentioned paper of L.R. Rabiner et al.

[0026] Performing the LPC analysis directly on the

coded signal affords a number of advantages in terms of processing power requirements. In particular, there is no need of dedicating processing power to the reconstruction of the LPC coefficients from the decoded signal: it is sufficient to extract them from the relevant information included in the coded speech signal, which is available on the same board. Besides the greater processing simplicity, also a reduction to at least of one fifth or even less of the information amount to be processed is achieved: indeed, at most 244 bits are to be processed to obtain the LPC coefficients from the coded signal, whereas 1280 bits are to be processed when the linear signal is used.

[0027] Under such conditions, the same processor used on the TRAU board for performing all TRAU functions and for managing the so-called tandem free operation (i.e. for dispensing with the transcoding in case of communication between two mobile terminals), for a plurality of speech channels in parallel (for instance 12 channels), can perform in real time, for the same channels, also the voice activity detection by exploiting both the spectral and the energy information, and the subsequent VQE. The resulting detection is more accurate than when only the energy information is exploited and hence also the noise suppression operation is more accurate.

[0028] It is to be appreciated that, according to the existing GSM standards, the LPC information in the coded signal is updated every 5 ms. The energy information is computed on the same interval of 5 ms and then the two contributions are jointly processed to take a decision on the nature of audio segment. This is denoted by the presence of said Joint Processing Means 43 and 45. For voice quality enhancement the high rate of decisions available at the output of said Hard decision Unit 43 is not necessary and has often a negative impact on audition. Therefore these "hard" decisions are softened through a smoothing process which aims to redefine as "voice" eventual isolated segments of noise among a group of segments of voice and to redefine as "noise" eventual isolated segments of voice among a group of segments of noise. This is the aim of Soft Decision Unit 45 sited immediately after said Hard Decision Unit 43.

[0029] It is clear that the above description is given only by way of non-limiting example and that variations and modifications are possible without departing from the scope of the invention. In particular, even if reference has been made to a TRAU unit in a GSM system, what has been said can be applied also to mobile communication systems operating according to other standards. In such case, the energy analysis and spectral analysis should be adapted to the specific requirements of that system. The invention could be used also in other radio communication signals in which a digital coding of speech is adopted and the digitally coded speech signals include spectral information. Moreover, also the energy analysis could be performed on the coded signal, as disclosed in the above-mentioned EP-A 1017042.

Claims

1. A method of detecting voice activity in a received speech signal in a radio communication system in which speech signals are transmitted in digitally coded form, and a signal representative of the presence or absence of voice activity is generated by submitting the received speech signals to an energetic analysis and a spectral analysis, **characterised in that** said spectral analysis is performed directly on coded speech signals.

2. A method as claimed in claim 1, wherein spectral information in said coded signals is periodically updated, **characterised in that** speech signal analysis comprises a joint processing of said spectral information with energy information, and generating said signal representative of the presence or absence of voice activity from said joint processing through a softening decision step adapted to smooth the decision's rate.

3. A method as claimed in claim 1 or 2, **characterised in that** said radio communication system is a mobile communication system and said signal representative of the presence or absence of voice activity is used to drive a noise reduction feature in a voice quality enhancement operation performed on signals propagating in uplink direction.

4. A voice activity detector for detecting voice activity in a received speech signal in a radio communication system in which digitally coded speech signals are transmitted, the detector (40) comprising means (41, 42, 43, 45) for performing an energetic analysis and a spectral analysis on the received speech signal and for generating a signal representative of the presence or absence of voice activity based upon the results of said analyses, **characterised in that** said spectral analysis means (41) are connected directly with a detector input (14) where said coded speech signals are present.

5. A voice activity detector as claimed in claim 4, wherein spectral information in said coded signal are periodically updated, **characterised in that** for a joint processing means (43, 45) of both spectral and energetic information are connected among said spectral (42) and energetic (41) analysis means and the output of the voice activity detector.

6. A voice activity detector as claimed in claim 5, **characterised in that** said Joint Processing Means are including:

   • an Hard Decision Unit (43) connected to said means (42) for performing an energetic analysis and to said means (41) for performing a

spectral analysis, adapted to joint processing the two input segments and to output an hard noise or voice decision;

- a Soft decision Unit (45) connected at the output of said Hard decision Unit (43) and adapted to perform a smoothing process in order to redefine as "voice" eventual isolated segments of noise among a group of segments of voice and to redefine as "noise" eventual isolated segments of voice among a group of segments of noise.

7. A voice activity detector as claimed in claim 4 to 6 for use in a mobile communication system, **characterised in that** said detector (40) is located upstream of means (50) performing a voice quality enhancement in the upstream direction of communication, and said signal representative of the presence or absence of voice activity drives noise reduction means in said means (50) performing voice quality enhancement.

8. A voice activity detector as claimed in claim 7, **characterised in that** said detector (40) is part, together with said voice quality enhancement means (50), of a unit (2) performing an adaptation of the uplink directed speech signals to the requirements of the control part of the mobile system and possibly of a fixed network and an adaptation of the downlink directed speech signals to the requirements of the radio part of the mobile system.

9. A voice activity detector as claimed in claim 8, **characterised in that** it is implemented by the same processor that would be provided for performing speech signal adaptation and voice quality enhancement in parallel for a plurality of speech channels.

10. A method of voice quality enhancement in a mobile communication system, in which a voice quality enhancement including a noise reduction operation is performed at least for speech signals propagating in uplink direction, **characterised in that** said noise reduction operation is driven by a signal representative of the presence or absence of voice activity generated by a method of voice activity detection as claimed in any of claims 1 to 3 and/or by a voice activity detector as claimed in any of claims 4 to 8.
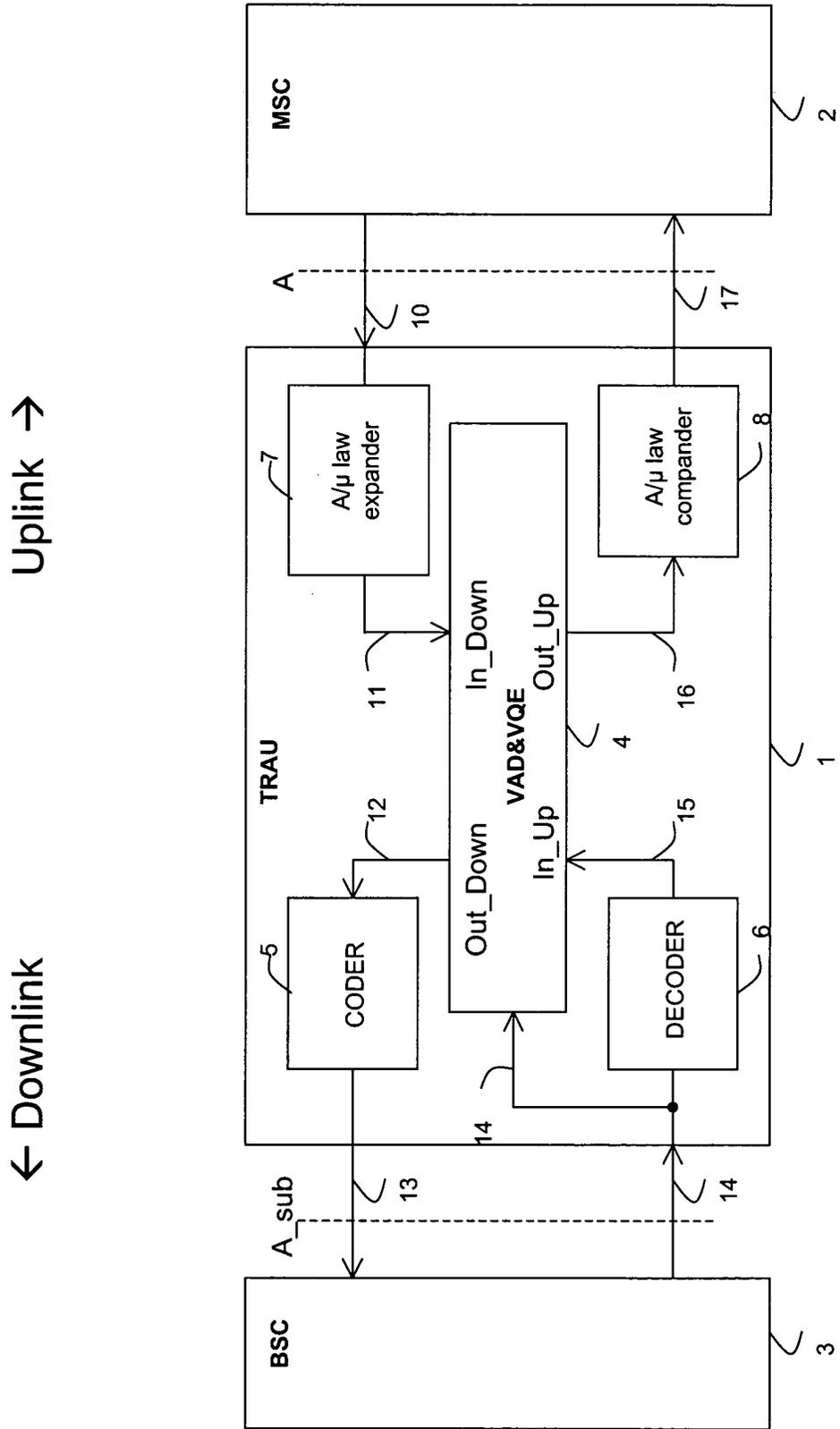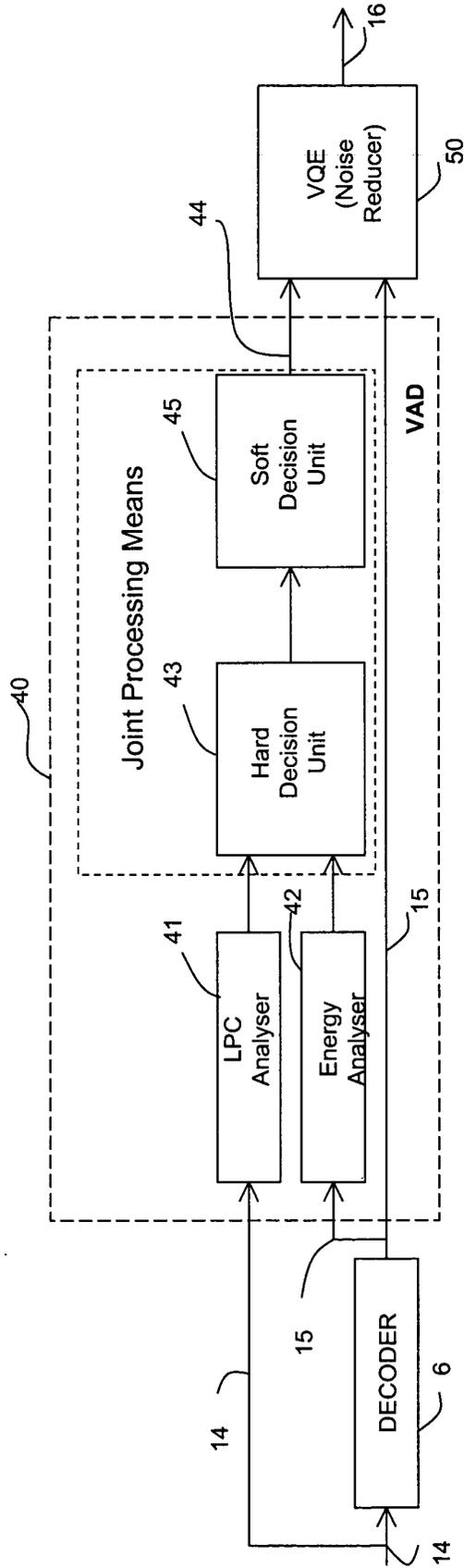
Uplink →

← Downlink

FIG. 1

FIG. 2

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 04 42 5031

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
D,Y	EP 1 017 042 A (AT & T CORP) 5 July 2000 (2000-07-05) * paragraphs [0018], [0021]; figure 1 * -----	1-10	G10L11/02 G10L19/00
Y	US 5 732 390 A (NISHIGUCHI MASAYUKI ET AL) 24 March 1998 (1998-03-24) * figures 1,2 * * column 3, line 40 - column 4, line 60 * -----	1-10	
A	US 5 579 435 A (JANSSON KLAS F J) 26 November 1996 (1996-11-26) * figure 2 * * column 3, line 19 - column 4, line 43 * -----	1,10	

TECHNICAL FIELDS SEARCHED (Int.Cl.7)

G10L

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	19 May 2004	Krembel, L

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

....

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.** EP 04 42 5031

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

19-05-2004

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 1017042	A	05-07-2000	US	5657422 A	12-08-1997
			EP	1017042 A1	05-07-2000
			CA	2138818 A1	29-07-1995
			CN	1132988 A	09-10-1996
			DE	69518174 D1	07-09-2000
			DE	69518174 T2	31-05-2001
			EP	0665530 A1	02-08-1995
			RU	2151430 C1	20-06-2000
US 5732390	A	24-03-1998	JP	7074709 A	17-03-1995
US 5579435	A	26-11-1996	SE	501981 C2	03-07-1995
			AU	672934 B2	17-10-1996
			AU	8118694 A	23-05-1995
			CA	2151398 A1	11-05-1995
			CN	1116011 A ,B	31-01-1996
			DE	69416668 D1	01-04-1999
			DE	69416668 T2	22-07-1999
			DK	677202 T3	27-09-1999
			EP	0677202 A1	18-10-1995
			ES	2131295 T3	16-07-1999
			FI	953261 A	30-06-1995
			GR	3029730 T3	30-06-1999
			HK	1014070 A1	14-07-2000
			JP	8505715 T	18-06-1996
			NZ	275905 A	26-11-1996
			SE	9303613 A	03-05-1995
			WO	9512879 A1	11-05-1995
			SG	47396 A1	17-04-1998

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82