

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
23 December 2004 (23.12.2004)

PCT

(10) International Publication Number
WO 2004/111995 A1

(51) International Patent Classification⁷: **G10L 11/02**,
H04R 3/00, 1/40

(21) International Application Number:
PCT/EP2004/051059

(22) International Filing Date: 8 June 2004 (08.06.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
03445076.7 17 June 2003 (17.06.2003) EP
60/480,876 24 June 2003 (24.06.2003) US

(71) Applicant (for all designated States except US): **SONY
ERICSSON MOBILE COMMUNICATIONS AB**
[SE/SE]; S-221 88 Lund (SE).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **GUSTAVSSON, Ste-
fan** [SE/SE]; Norra Stenbocksgatan 46, S-254 43 Helsing-
borg (SE).

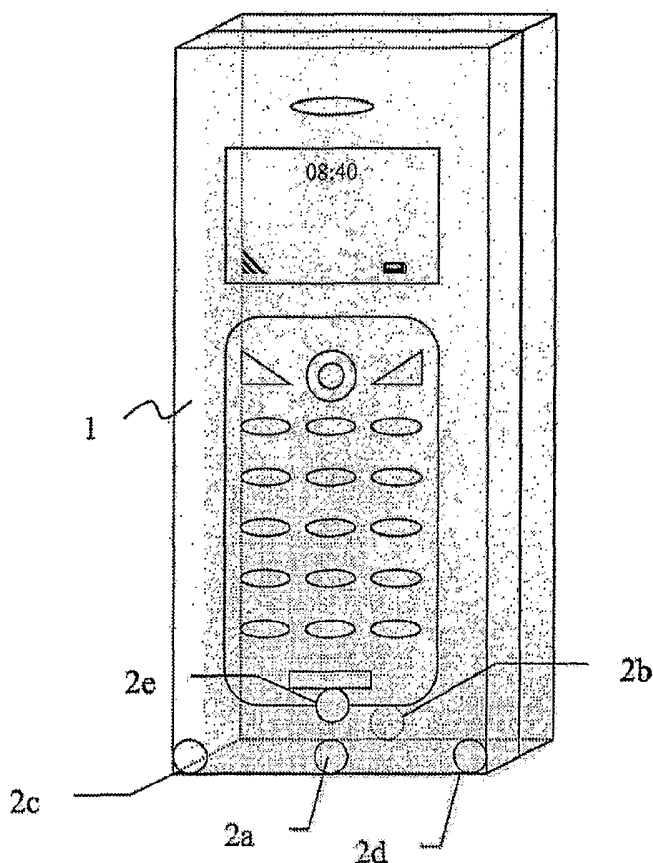
(74) Agents: **AKERMAN, Marten** et al.; Albihns Malmö AB,
P.O. Box 4289, S-203 14 Malmö (SE).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,

[Continued on next page]

(54) Title: DEVICE AND METHOD FOR VOICE ACTIVITY DETECTION



(57) Abstract: The invention relates to a device, a mobile apparatus incorporating the device, and accessory therefor and a method for voice activity detection, particularly in a mobile telephone, using the directional sensitivity of a microphone system and exploiting the knowledge about the voice source's orientation in space. The device comprises a sound signal analyser arranged to determine whether a sound signal comprises speech. According to the invention, the device further comprises a microphone system (2a, 2b, 2c, 2d, 2e) arranged to discriminate sounds emanating from sources located in different directions from the microphone system, so that sounds only emanating from a range of directions are included as signals possibly containing speech.



GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

Declaration under Rule 4.17:

- *of inventorship (Rule 4.17(iv)) for US only*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DEVICE AND METHOD FOR VOICE ACTIVITY DETECTION

Field of the invention

The present invention relates to a device, a mobile apparatus incorporating
5 the device, an accessory therefor and a method for voice activity detection,
particularly in a mobile telephone, using the directional sensitivity of a microphone
system and exploiting the knowledge about the voice source's orientation in space.
The device assists the existing voice activity detection to achieve higher sensitivity
and requiring less processor power.

10

State of the art

Voice activity detectors are used e.g. in mobile phones to enhance the
performance in certain situations. The most common way to construct a voice
activity detector is to look at the levels of the sub-bands of the incoming signal.
15 Then the background noise level and the speech level are estimated and compared
with a threshold to determine whether speech is present or not. An example of a
voice activity detector is disclosed in U.S. patent 6,427,134.

For instance in noisy environments it is hard to make a uniform parameter
set-up for the voice activity detector. Therefore several voice activity detectors are
20 needed, trimmed to the specific cases. For example in some modules you need to be
sure that if there is speech it should be detected (echo canceller), but in other cases
it is better to indicate no speech if the signal to noise ratio level is too low. The
plurality of voice activity detectors put a load on the digital signal processors that
have to take care of performing the various voice activity detection algorithms.

25

Summary of the invention

An object of the present invention is to complement existing voice activity
detection taking into account the direction of the source of the sound.

In a first aspect, the invention provides a device for voice activity detection
30 comprising a sound signal analyser arranged to determine whether a sound signal
comprises speech.

According to the invention, the device further comprises
a microphone system arranged to discriminate sounds emanating from sources
located in different directions from the microphone system, so that sounds only
35 emanating from a range of directions are included as signals possibly containing
speech.

Suitably, the range of directions is directed in the direction of an intended
user's mouth.

In one embodiment, the microphone system comprises two microphone

elements separated a distance and located on a line directed in the direction of an intended user's mouth.

The range of directions may be defined as all sounds falling inside a cone with a cone angle α , wherein $10^\circ < \alpha < 30^\circ$, and preferably, α is approximately 25° .

5 In another embodiment, the microphone system comprises three microphone elements separated a distance and located in a plane directed in the direction of an intended user's mouth.

Suitably, two of said three microphone elements are separated a distance and located on a line directed perpendicular to the direction of an intended user's mouth.

10 In another embodiment, the microphone system comprises four microphone elements located such that the fourth microphone is not located in the same plane as the three others.

The microphone elements may be directional with a pattern having maximal sensitivity in the direction of an intended user's mouth.

15 In still a further embodiment, the microphone system comprises one directional microphone element together with one or more other microphone elements to remove the uncertainty in the direction of the sound source. The directional microphone element may be used to measure the sound pressure level relative to the other microphone element.

20 In a second aspect, the invention provides a mobile apparatus comprising a device as mentioned above.

Suitably, the microphone elements are located at the lower edge of the apparatus.

25 In one embodiment, a plurality of microphone elements are located at the lower edge of the apparatus and at least one further microphone element is located at a distance from the lower edge.

The mobile apparatus may be a mobile radio terminal, e.g. a mobile telephone, a pager, a communicator, an electric organiser or a smartphone.

30 In a third aspect, the invention provides an accessory for a mobile apparatus comprising a microphone system as mentioned above.

Suitably, the direction of the range of directions is adjustable.

The accessory may be a hands-free kit or a telephone conference microphone.

35 In a fourth aspect, the invention provides a method for voice activity detection, including the steps of:
receiving sound signals from a microphone system arranged to discriminate sounds emanating from sources located in different directions from the microphone system;
determining the direction of the sound source causing the sound signals;
if the sounds emanate from a first range of directions, further analyse

the sound to determine whether the sound signal comprises speech;
but if the sounds emanate from a second, different range of directions decide that
the sound signal does not comprise speech.

Suitably, the first range of directions is directed in the direction of an
5 intended user's mouth.

The first range of directions may be defined as all sounds falling inside cone
with a cone angle α , wherein $10^\circ < \alpha < 30^\circ$, and preferably α is approximately 25° .

In one embodiment, the microphone system comprises at least two
microphone elements located at a distance from each other and located on a line
10 directed in the direction of an intended user's mouth, said two microphone elements
being separated a distance d , wherein the direction to the sound source θ is
calculated as

$$\theta = \arccos \frac{\Delta t \cdot v}{2 \cdot d}$$

where

15 Δt is the time difference between the sounds from the two microphone elements,
 v is the velocity of sound.

In another embodiment, one directional microphone element is used together
with one or more other microphone elements to remove the uncertainty in the
direction of the sound source.

20 The directional microphone element may be used to measure the sound
pressure level relative to the other microphone element.

The invention is defined in the attached independent claims 1, 12, 16, and 20,
while preferred embodiments are set forth in the dependent claims.

25 Brief description of the drawings

The invention will be described below in greater detail with reference to the
accompanying drawings, in which:

fig. 1 is a perspective view of a mobile phone incorporating the present
invention, and

30 fig. 2 is a schematic drawing of the receiving angle of an embodiment of the
present invention.

Detailed description of preferred embodiments

As mentioned briefly in the introduction, many signal processing algorithms,
35 such as echo cancellation and background noise synthesis, used in phones and
hands-free kits are based on the fact that the user is speaking or not. For example
the speech codec is active when the near-end user is speaking and the background
synthesis is active when the near-end user is silent. All these algorithms need good

voice activity detectors (VAD) to perform well. An error in the detection can result in artefacts or malfunctions caused by divergence of the algorithms or other problems.

Existing voice activity detectors are directed to determine whether speech is present or not in a sound signal. However, in fact not all speech is interesting or relevant, but only the user's speech. All other speech, e.g. in a noisy environment with several persons speaking, could be ignored and regarded as just noise.

The present inventor has realised that a microphone system having some kind of directional sensitivity could be used to discriminate sound emanating from different sources located in different directions. Sound not emanating from the user can be declared as non-speech, and those signals do not have to be analysed with the conventional voice activity detectors.

The existing voice activity detectors may be conventional and are only referred to as a sound signal analyser in this application.

Generally, a microphone system having some kind of directional sensitivity can be used. Fig. 1 shows an example with at least two separate microphone elements.

A general mobile telephone is indicated at 1. The invention is equally applicable to other devices such as mobile radio terminals, pagers, communicators, electric organisers or smartphones. The common feature is that voice activity detection is employed, e.g. in connection with communicating speech or receiving voice commands by means of speech recognition.

In the simplest version, the microphone system comprises two microphones 2a and 2b. Suitably, they are located on a line directed in the calculated direction of an intended user's mouth. Suitably, the microphone elements are located at the lower edge of the mobile apparatus 1.

Fig. 2 shows a schematic diagram of the calculation of the direction of the sound source, typically the user's mouth 3. In the case of two microphones, only the angle to the line on which the microphone elements are located can be determined. In other words, the direction of the sound source is on a cone with a cone angle θ . To calculate the angle θ , first a cross-correlation between the two signals from the microphones 2a and 2b is made. The maximum indicates the time difference Δt between the two microphones 2a and 2b. The distance between the two microphones 2a and 2b is e.g. 20 millimetres. The angle θ is calculated as

$$\theta = \arccos \frac{\Delta t \cdot v}{2 \cdot d}$$

Note that arccos is only defined for arguments between -1 and 1. If the time

difference is negative, this means that the angle is greater than 90° and the sound emanates from behind the apparatus.

Suitably, the device is adapted to determine that all sounds with an angle θ less than a fixed angle α are emanating from the user. The threshold angle α may be
5 set within a range of e.g. 10° to 30° , suitably at 25° .

In the case of three microphones, the direction of the sound source can be further determined to be at two points (e.g. on the above cone). The three microphone elements are suitably located in a plane directed in the general direction of the user's mouth. In fig. 1 microphone elements 2b, 2c and 2d are a possible set-
10 up. The two microphone elements 2c and 2d at the front are located on a line perpendicular to the direction of the user's mouth, while the third microphone element 2b is located at the rear side.

In the case of four microphones (or more) detection of all direction angles may be calculated, provided that four microphone elements are located such that the
15 fourth microphone is not located in the same plane as the three others, e.g. on a tetrahedron. A possible set-up is two microphone elements 2c and 2d at the front on the lower edge, while a third microphone element 2b is located at the rear side, and a fourth microphone element 2e is located at the front at a distance from the lower edge.

20 A similar microphone arrangement may be used in an accessory to a mobile apparatus, such as a hands-free kit or a telephone conference microphone system intended to be placed on a table. Apart from the microphone elements the logic circuitry may be located in the main/mobile apparatus. In this case the reception angle of the microphone system can be adjustable. This is useful e.g. when the
25 microphone system is placed in a car, where the user can be seated either in the driver's seat or in the passenger's seat or even both the driver and the passenger may be speakers during the same call. The adjustment of the reception angle can be achieved mechanically or electronically, for example by beam forming or adaptation of the directional sensitivity of the microphone system.

30 To further enhance the sensitivity of the microphone system, directional microphone elements with a pattern having a maximum sensitivity in the direction of the user's mouth could be used.

In a further embodiment, one directional microphone element is used together with one or two other microphone elements (that may be non-directional).
35 The directional microphone element is used to measure the sound pressure level relative to the other(s), thus removing the uncertainty in the direction of the sound source. Various combinations of directional microphone elements and non-directional microphone elements are possible.

The present invention leads to a voice activity detector having enhanced

performance. With the present invention only one voice activity detector may be necessary throughout the whole signal path. This will in turn reduce the computational complexity, decreasing the load on the digital signal processors as well as improving the performance. It is especially favourable in environments with
5 high background noise and noise with similar spectral properties as speech.

A person skilled in the art will realise that the invention may be realised with various combinations of hardware and software. The scope of the invention is only limited by the claims below.

CLAIMS

1. A device for voice activity detection comprising a sound signal analyser arranged to determine whether a sound signal comprises speech, **characterised** by
5 a microphone system (2a, 2b, 2c, 2d, 2e) arranged to discriminate sounds emanating from sources located in different directions from the microphone system, so that sounds only emanating from a range of directions are included as signals possibly containing speech.
- 10 2. A device according to claim 1, **characterised** in that the range of directions is directed in the direction of an intended user's mouth (3).
3. A device according to claim 2, **characterised** in that the microphone system comprises two microphone elements (2a, 2b) separated a distance and located on
15 a line directed in the direction of an intended user's mouth (3).
4. A device according to claim 3, **characterised** in that the range of directions is defined as all sounds falling inside a cone with a cone angle α , wherein
20 $10^\circ < \alpha < 30^\circ$.
5. A device according to claim 3, **characterised** in that α is approximately 25° .
6. A device according to claim 2, **characterised** in that the microphone system comprises three microphone elements (2b, 2c, 2d) separated a distance and
25 located in a plane directed in the direction of an intended user's mouth (3).
7. A device according to claim 6, **characterised** in that two (2c, 2d) of said three microphone elements are separated a distance and located on a line directed perpendicular to the direction of an intended user's mouth (3).
30
8. A device according to claim 2, **characterised** in that the microphone system comprises four microphone elements (2b, 2c, 2d, 2e), located such that the fourth microphone (2e) is not located in the same plane as the three others (2b, 2c, 2d).
35
9. A device according to any one of claims 1 to 8, **characterised** in that the microphone elements (2a, 2b, 2c, 2d, 2e) are directional with a pattern having maximal sensitivity in the direction of an intended user's mouth (3).

10. A device according to claim 1, **characterised** in that the microphone system comprises one directional microphone element together with one or more other microphone elements adapted to remove the uncertainty in the direction of the sound source.
- 5 11. A device according to claims 10, **characterised** in that the directional microphone element is adapted to measure the sound pressure level relative to the other microphone element.
- 10 12. A mobile apparatus, **characterised** in that it comprises a device as defined in any one of claims 1 to 11.
13. A mobile apparatus according to claim 12, **characterised** in that the microphone elements (2a, 2b, 2c, 2d) are located at the lower edge of the apparatus.
- 15 14. A mobile apparatus according to claim 12, **characterised** in that a plurality of microphone elements (2a, 2b, 2c, 2d) are located at the lower edge of the apparatus and at least one further microphone element (2e) is located at a distance from the lower edge.
- 20 15. A mobile apparatus according to any one of claims 12 to 14, **characterised** in that it is a mobile radio terminal, e.g. a mobile telephone (1), a pager, a communicator, an electric organiser or a smartphone.
- 25 16. An accessory for a mobile apparatus, **characterised** in that it comprises a microphone system (2a, 2b, 2c, 2d, 2e) as defined in any one of claims 1 to 11.
17. An accessory according to claim 16, **characterised** in that the direction of the range of directions is adjustable.
- 30 18. An accessory according to claim 16 or 17, **characterised** in that it is a hands-free kit.
- 35 19. An accessory according to claim 16 or 17, **characterised** in that it is a telephone conference microphone.
20. A method for voice activity detection, **characterised** by the steps of: receiving sound signals from a microphone system (2a, 2b, 2c, 2d, 2e) arranged

- to discriminate sounds emanating from sources located in different directions from the microphone system;
determining the direction of the sound source causing the sound signals;
if the sounds emanate from a first range of directions, further analyse
5 the sound to determine whether the sound signal comprises speech;
but if the sounds emanate from a second, different range of directions decide that the sound signal does not comprise speech.
21. A method according to claim 20, **characterised** in that the first range of
10 directions is directed in the direction of an intended user's mouth (3).
22. A method according to claims 21, **characterised** in that the first range of directions is defined as all sounds falling inside a cone with a cone angle α , wherein $10^\circ < \alpha < 30^\circ$.
15
23. A method according to claims 22, **characterised** in that α is approximately 25° .
24. A method according to any one of claims 22 or 23, **characterised** in that the
20 microphone system comprises at least two microphone elements (2a, 2b) located at a distance from each other and located on a line directed in the direction of an intended user's mouth (3), said two microphone elements being separated a distance d , wherein the direction to the sound source θ is calculated as
$$\theta = \arccos \frac{\Delta t \cdot v}{2 \cdot d}$$

25 where
 Δt is the time difference between the sounds from the two microphone elements,
 v is the velocity of sound.
25. A method according to claims 20, **characterised** in that one directional
30 microphone element is used together with one or more other microphone elements to remove the uncertainty in the direction of the sound source.
26. A method according to claims 25, **characterised** in that the directional
35 microphone element is used to measure the sound pressure level relative to the other microphone element.

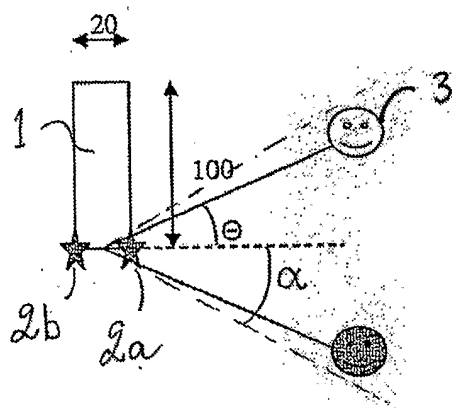
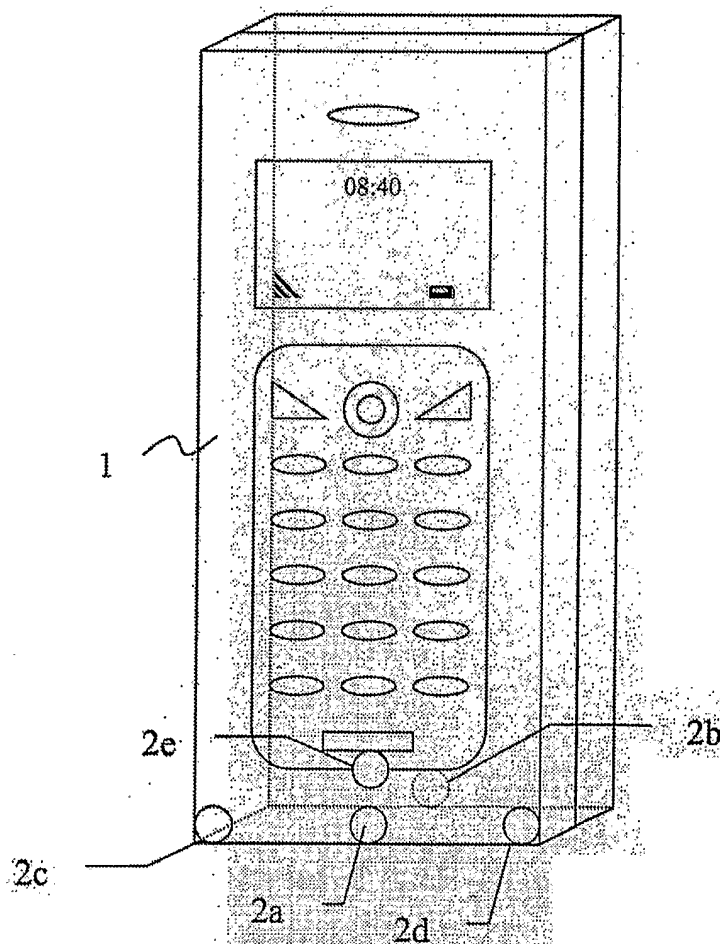


FIG 2

FIG 1



INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP2004/051059

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L11/02 H04R3/00 H04R1/40

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L H04R

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2003/027600 A1 (ORAINARA SOOTORN ET AL) 6 February 2003 (2003-02-06) paragraph '0006! paragraph '0017! paragraph '0022!	1, 2, 12, 15
Y	idem	3, 6-11, 13, 14, 16-19
A		20
Y	US 2002/009203 A1 (ERTEN GAMZE) 24 January 2002 (2002-01-24) figures 2, 4 paragraph '0009! paragraphs '0044!, '0045!	3
Y	figures 9, 12	13, 16-19
Y	paragraph '0070!; figure 4	9-11
	----- -/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

15 September 2004

Date of mailing of the international search report

20/10/2004

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Krembel, L

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP2004/051059

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 1 206 161 A (SONY INTERNAT EUROP GMBH) 15 May 2002 (2002-05-15) abstract sentence 40 - sentence 44 paragraph '0021! paragraphs '0027!, '0028! figures 2,3 -----	6-8, 14

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP2004/051059

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 2003027600	A1	06-02-2003	NONE	
US 2002009203	A1	24-01-2002	AU 5121301 A	15-10-2001
			CA 2404071 A1	11-10-2001
			CN 1436436 T	13-08-2003
			EP 1295507 A2	26-03-2003
			JP 2003530051 T	07-10-2003
			WO 0176319 A2	11-10-2001
EP 1206161	A	15-05-2002	EP 1206161 A1	15-05-2002