(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 July 2008 (10.07.2008)

PCT

(10) International Publication Number
WO 2008/082308 A1

(51) International Patent Classification:
H04R 1/28 (2006.01) *H04N 7/15* (2006.01)

(21) International Application Number:
PCT/NO2007/000455

(22) International Filing Date:
20 December 2007 (20.12.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
NO 20066067 29 December 2006 (29.12.2006) NO

(71) Applicant *(for all designated States except US)*: **TANDBERG TELECOM AS** [NO/NO]; Philip Pedersens vei 22, N-1366 Lysaker (NO).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ENSTAD, Gisle** [NO/NO]; Haugjordet 17, N-1336 Sandvika (NO). **MARTON, Trygve Frederik** [NO/NO]; Hoff Terrasse 19, N-0275 Oslo (NO). **NIELSEN, Johan Ludvig** [—/NO]; Oberst Rodes vei 26, N-1152 Oslo (NO).

(74) Agent: **OSLO PATENTKONTOR AS**; Postboks 7007 M, N-0306 Oslo (NO).

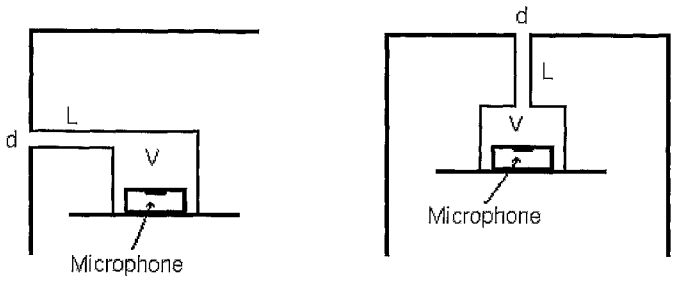
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN,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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, 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, IS, IT, LT, LU, LV, MC, MT, 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).

Declaration under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:
— *with international search report*

(54) **Title:** MICROPHONE FOR AUDIO SOURCE TRACKING



(57) **Abstract:** The present invention discloses an arrangement utilizes a certain microphone assembly for facilitating audio source tracking systems in communication systems. It can be applied to both single and array microphones. The principal idea is to enhance the sound level acoustically in the critical high frequency range, thereby increasing the effective signal-to-noise ratio both for sound pickup and localization algorithms. This is done by enclosing the microphone into a channel or a small cavity (a Helmholtz-resonator), thereby introducing a high-frequency response peak (resonance), fairly broad-band.

WO 2008/082308 A1

MICROPHONE FOR AUDIO SOURCE TRACKING

Field of the invention

The present invention relates to sound source localization, in particular a microphone arrangement particularly configured for capturing audio signals to be used in signal source localization.

Background of the invention

Signal localization is used in several applications. The most known application is perhaps TV program production. In e.g. debate programs, it is important for the viewer's experience and intelligibility that the active camera is pointing at, and preferably zooming on, the current speaker. However, this has traditionally been handled manually by a producer. In other applications where cameras and microphones are capturing the view and sound of a number of people, it might be impossible or undesirable to have a dedicated person to control the performance.

One example of such application is automatic camera pointing in video conferencing systems. A typical situation at an end-point in a video conference call is a meeting room with a number of participants sitting around a table watching the display device of the end-point, while a camera positioned near the display device is capturing a view of the meeting room. If there are many participants in the room, it may be difficult for those who are watching the view of the meeting room at a far end side to determine the speaker or to follow the speaker's arguing. Thus, it would be preferable to localize the active speaker in the room, and automatically point and/or zoom the camera onto that participant. Automatically orienting and zooming of a camera given a certain position within reach of the camera, is well known in the art, and will not be discussed in detail. The problem is to provide a sufficiently accurate

localisation of the active speaker, both in space and in time, in order to allow acceptable automatic video conference production.

Known audio source localization arrangements are using a plurality of spatially spaced microphones, and are often based on the determination of a delay difference between the signals at the outputs of the receivers. If the positions of the microphones and a delay difference between the propagation paths between the source and the different microphone are known, the position of the source can be determined.

One example of an audio source localisation is shown in U.S. patent number 5,778,082. This patent teaches a method and a system using a pair of spatially separated microphones to obtain the direction or location of an audio source. By detecting the beginning of the respective signals of the microphones representing the sound of the same audio source, the time delay between the audio signals may be determined, and the distance and direction to the audio source may be calculated.

If three microphones are used, it becomes possible to determine a position of the source in a 2-D plane. If more than three microphones, not placed in a single plane, are used, it is possible to determine the position of a source in three dimensions. A common assembly is the placement of one array of microphones in the horizontal direction below the camera, and one single microphone above the camera. This allows both horizontal and vertical source localization. The microphone mounted above the camera may be very dominant visually, and are exposed to damages and may introduce extra manufacturing costs. A solution were a microphone is integrated in top of the camera itself is therefore preferable. This allows for vertical localization

of a source without having a microphone mounted on e.g. a rod above the camera. The microphone is invisible and well protected, and provides possibilities for less intrusive design, more visually pleasing. As indicated above, this preference has some disadvantages due to smaller distances between the microphones, implying less accuracy in the signal source tracking.

In addition, sound quality is an important issue for source tracking applications, as good signal quality is also necessary for tracking accuracy. When outputs from several microphones are combined, precise and repeatable frequency response is required, both amplitude and phase, and wide band. Matching requirements can be lower than 1dB and a few degrees phase. This is not found in normal transducer production, not even with high cost microphones. Therefore, manufacturers have to match transducers by measurement and sorting, or product designers must incorporate some form of measurement and calibration of individual microphones. Both alternatives are costly.

Another recently emerged alternative is to utilize new microphone technology, MEMS (Micro Electro Mechanical Systems). MEMS microphones are produced using silicon wafer technology, and the process gives microphones with variation in phase response that is significantly less than regular ECM (Electret Condenser Microphone) microphones. They are therefore well suited for applications where good phase matching of microphones is required.

Microphone self-noise is, however, a real problem, especially with cheap ECMs. MEMS microphones have even higher self noise than standard ECMs, which is a problem for sound pickup and localization systems, especially at high frequencies.

Background noise in rooms typically has decreasing power with increasing frequency, while many cheap microphone

types have a constant or increasing self-noise power with increasing frequency. Speech signals have very low power in the high frequencies, but the high frequency content is still important for natural sounding speech recordings and also provides very effective cues for source localization algorithms. In the high frequencies, the microphone self-noise is the dominant noise contributor, and this limits the signal to noise ratio (SNR) when capturing speech in rooms. This is especially true when microphones cannot be employed close to the persons talking, and limits the potential for localization and tracking algorithms. Analysis algorithms are disturbed because high frequency information in speech is masked by microphone self-noise.

Summary of the invention

The present invention discloses an audio source tracking arrangement for determining a position of a source creating a sound comprising an audio signal processing module configured to determine the position by means of a number of audio signals originating from the sound respectively captured by said number of microphones and one or more microphone housings respectively encapsulating at least one of said number of microphones including a cavity in which at least one microphone is localized, an aperture on a surface of the encapsulation device and a channel extending from the cavity to the aperture.

Brief description of the drawings

In order to make the invention more readily understandable, the discussion that follows will refer to the accompanying drawings,

Figure 1 shows a cross section of microphone housing according to one embodiment of the present invention,

Figure 2 and 3 shows an example assembly with a microphone housing, a PCB card and a camera montage,

Figure 4 is a plot of an example of a resulting frequency response when using the present invention.

Best modes of carrying out the invention

In the following, the present invention will be discussed by describing some example embodiments, and by referring to the accompanying drawings. However, people skilled in the art will realize other applications and modifications within the scope of the invention as defined in the enclosed independent claims.

The present invention utilizes a certain microphone assembly for facilitating audio source tracking systems in communication systems. It can be applied to both single and array microphones. The principal idea is to enhance the sound level acoustically in the critical high frequency range, thereby increasing the effective signal-to-noise ratio both for sound pickup and localization algorithms.

This is done by enclosing the microphone into a channel or a small cavity (a Helmholtz-resonator), thereby introducing a high-frequency response peak (resonance), fairly broad-band.

Figure 1 exemplifies this aspect of the present invention. As can be seen, the microphone element is embedded in an assembly principle. The figure shows a cross section of the principle. The assembly comprises a sound inlet channel with diameter d and length L, and a cavity with volume V enclosing the microphone. The sound inlet channel forms an acoustical resonator together with the cavity encapsulating the microphone. This gives an amplified acoustical response in a fairly broad frequency band, improving signal to microphone self-noise ratio in that band. The resonance

frequency results from interaction between sound inlet channel diameter d and length L, and the cavity volume V. The shorter channel length L, and smaller cavity volume V, the higher the resonance frequency. Increasing channel diameter d also increases resonance frequency. According to the Helmholtz theory, the resonance frequency (f_r) can be calculated as follows:

$$f_r = \frac{c}{2\pi}\sqrt{\frac{\pi(\frac{d}{2})^2}{VL}}$$

Figures 2-3 show a further aspect of the embodiment of the present invention. It includes a surface mount MEMS microphone encapsulated by a hard plastic housing on a small PCB card. This PCB card with microphone and plastic housing is mounted in a rigid plastic construction. Together these parts form an acoustical resonator system with an inlet sound channel, and a resonant cavity enclosing the microphone as described above. The assembly serves a multitude of functions. As already mentioned, the channel length and volume are adapted to give a resonance peak in the desired high frequency range. The assembly is constructed to ensure easy and secure mounting. The construction also serves to protect the microphone from physical impact and damage, as well as electro static discharge (ESD).

Mechanical protection of the element is preferably secured by making the housing sturdy and rugged out of a hard material. The material should be non-porous so as to minimize sound absorption. An elastomer cast with 80 Shore A hardness is a working compromise. It should be somewhat elastic to withstand varying stresses from the system above it, and to provide a good acoustic seal to ensure that there is no leakage of air from the cavity around the microphone.

The assembly should be acoustically sealed by designing it so that the outer plastic part, to which the microphone card with the microphone housing is mounted, exerts a small pressure on top of the microphone housing. Also, the sound inlet hole to the microphone housing is tightly fitted on to a tube on the surrounding plastic part, extending the sound inlet hole to the front of the outer plastic part.

In this embodiment, ESD protection is achieved by having an exposed conducting pad on the PCB card in front of the microphone, so that the conducting pad is closer to the sound inlet channel than the microphone element is. An ESD pulse through the sound inlet channel will hit the conducting pad, and not the microphone because the pulse will always go to ground the shortest possible way.

Any microphone element requiring sound wave entry from a single direction can be used in the present invention. However, a typical omni directional MEMS microphone is chosen in the described embodiment for repeatable phase response. The size of the element is in principle not important, but the size of the microphone itself, and its sound inlet, has an impact on the minimal size of the cavity enclosing the microphone.

The total free-field response of the microphone in its housing, is a convolution of the microphone response, the entry channel volume response, and the pressure-build up effect on the front of the assembly. A high frequency response peak sized and shaped by the mechanical design will invariably result.

For automatic camera control in video conferencing, a microphone array for speaker localization is needed. In one embodiment of the present invention, a small number of microphones, typically four, are mounted in connection with the camera. Typically, three microphones form a horizontal array below the camera, and one is placed above, but

integrated in the camera forming a vertical microphone pair with the middle microphone in the array below the camera. This allows both horizontal and vertical source localization. The microphone mounted above the camera is very dominant visually. According to this embodiment, the microphone is invisible and well protected, and provides possibilities for less intrusive design, more visually pleasing.

Figure 4 shows a plot of the resulting amplitude response of a microphone mounted in the solution described above, versus the response of the same microphone in open air. As can be seen a resonant peak gives from 3 to 7 dB acoustic amplification in the frequencies from about 8 kHz to 11 kHz. The shape and frequency range of the resonance can be changed by adjusting the mechanical design enclosing the microphone.

This influence on the frequency response is dominant compared to the effect of reflection and diffraction from nearby objects, and can therefore provide a response with less variation for changing angle from the sound source, which is advantageous.

For most other purposes than audio source tracking, the frequency response is preferred to be flat to avoid sound distortion. In this case, the sound captured by the microphone is not supposed to reach a human ear, but is instead used for detection. It optimizes clarity and improves the signal-to-noise ratio in the following analog-to-digital converter if the system is digital. Due to the noise and a relatively short distance between the microphone elements, it is important in sound source detection to have a strong detection signal relative to the noise floor. Thus, a peak in the frequency response in the high frequency area is in this case an advantage.

Patent claims

1. An audio source tracking arrangement, integrated in or connected to a video conference system, for determining a position of a source creating a sound, comprising at least an audio signal processing module configured to determine the position by means of a number of audio signals originating from the sound respectively captured by said number of microphones
characterized in

one or more microphone housings, respectively encapsulating at least one of said number of microphones, comprising a cavity in which the at least one microphone is localized, an aperture on a surface of the microphone housing and a channel extending from the cavity to the aperture, and where

said channel and said cavity are dimensioned to form a acoustical amplifier with a frequency response having one or more high frequency peaks in a frequency band of the sound.

2. The sound source tracking arrangement according to claim 1,
characterized in that said number of microphones are MEMS microphones.

3. The sound source tracking arrangement according to one of the claims 1 - 3,
characterized in that said number of microphones and the one or more microphone housings are integrated in a camera cover.

4. The sound source tracking arrangement according to one of the claims 1 - 4,
characterized in that the position defines

a reference point relative to which a camera and/or a processor adjusts a captured or viewed image.

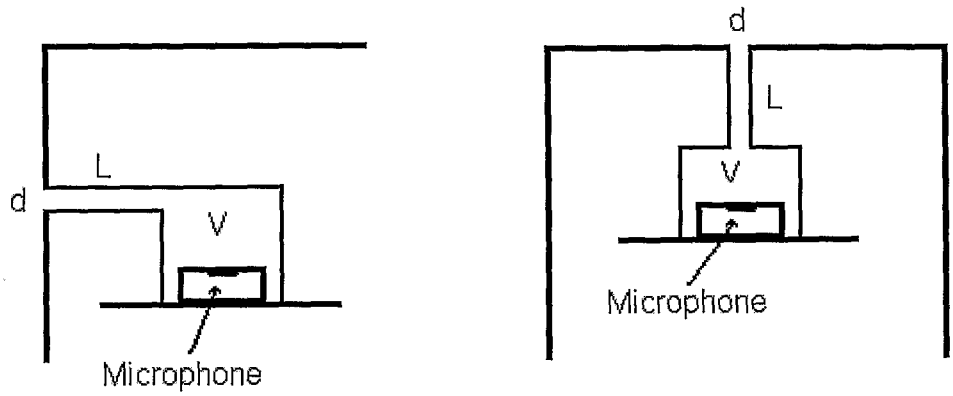


Figure 1

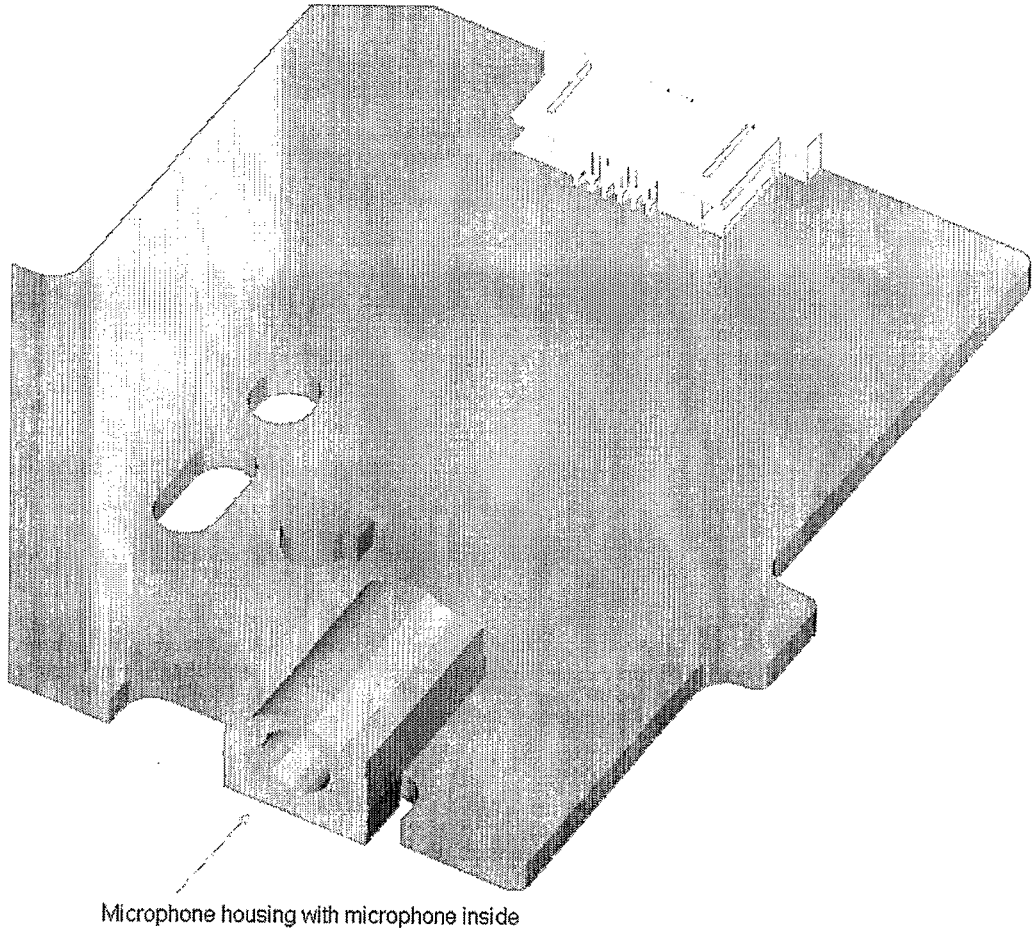


Figure 2

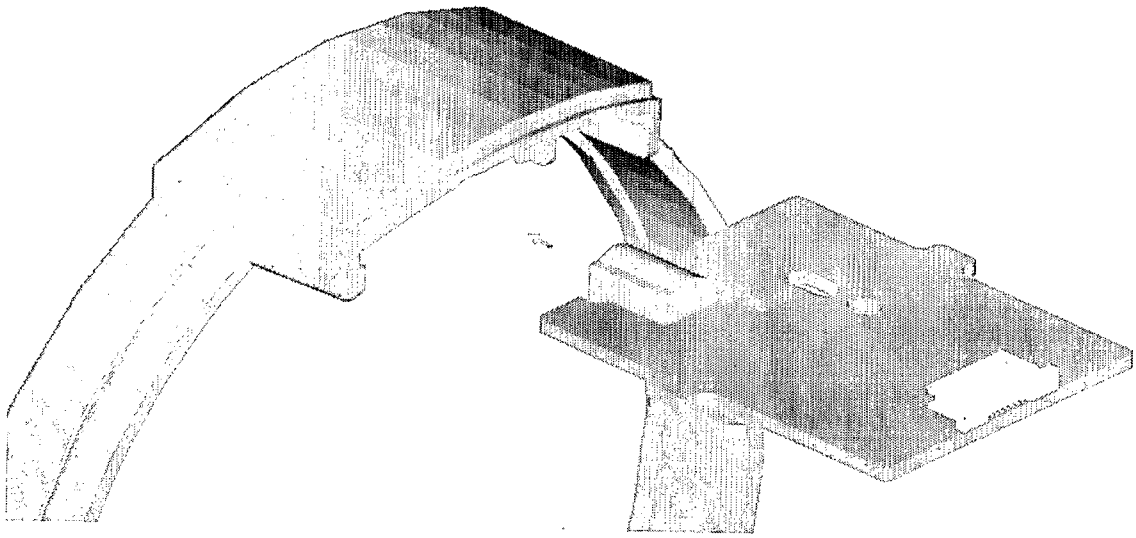


Figure 3

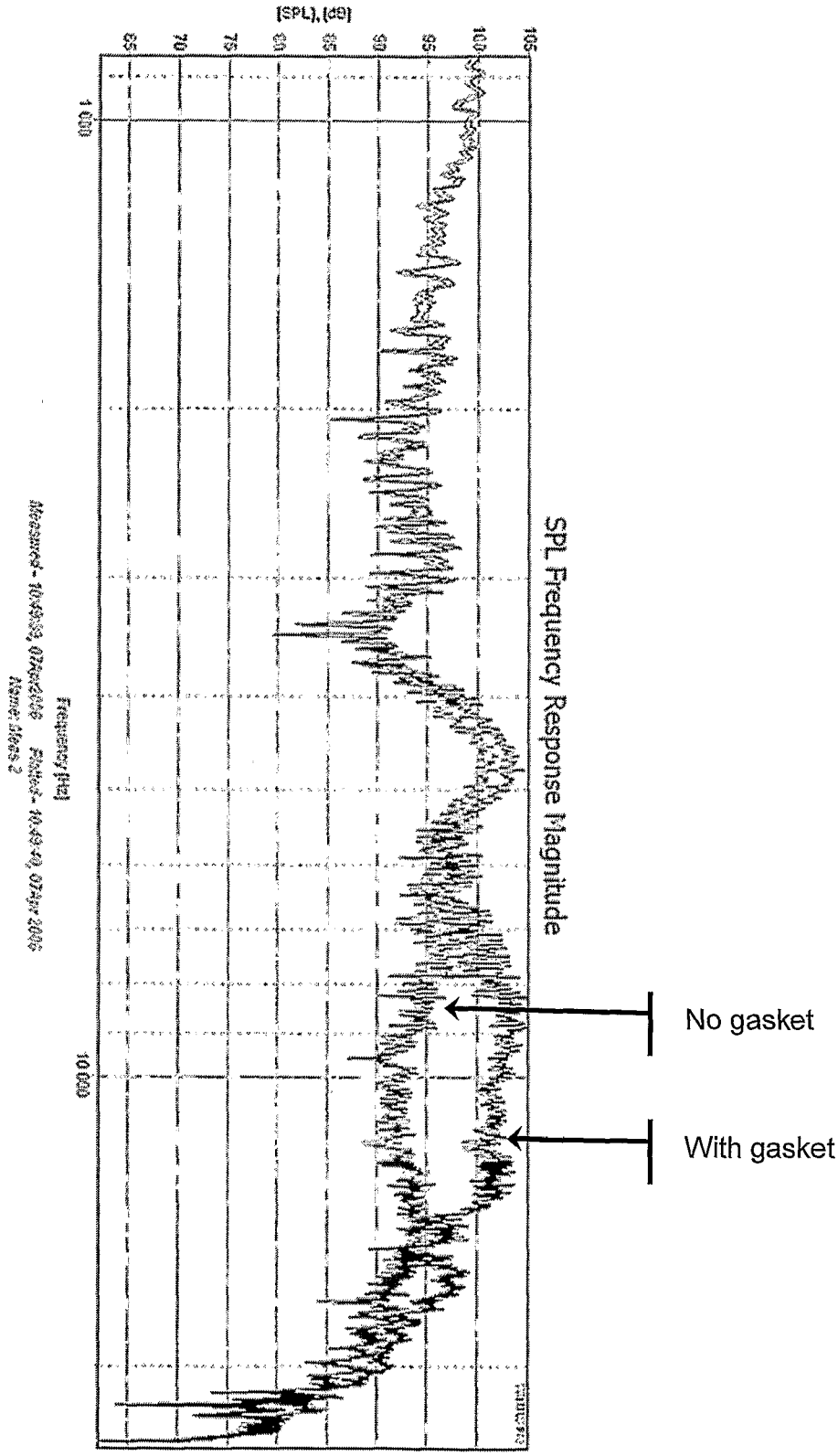


Figure 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NO2007/000455

A. CLASSIFICATION OF SUBJECT MATTER

IPC: see extra sheet
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: H04R, H04N

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

SE,DK,FI,NO classes as above

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

EPO-INTERNAL, WPI DATA, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GB 2234137 A (MESSERSCHMITT-BÖLKOW-BLOHM GMBH), 23 January 1991 (23.01.1991), the whole document ---	1-4
A	WANG, H ET AL: "Voice source localization for automatic camera pointing system in videoconferencing", Acoustics, Speech, and Signal Processing. 1997. ICASSP-97., 1997 IEEE International Conference on Munich, Germany 21 -24 April 1997, Los Alamitos, CA, USA. IEEE Comput. Soc, US, 1997-04-21, ISBN 0-8186-7919-0, vol 1, pages 187-190 ---	1-4
A	US 120030228025 A1 (HANNAH, E C), 11 December 2003 (11.12.2003) ---	1-4

[X] Further documents are listed in the continuation of Box C. [X] See patent family 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 application or patent 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	Date of mailing of the international search report
3 April 2008	11 -04- 2008
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Henrik Andersson /itw Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NO2007/000455

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0836324 A2 (PICTURETEL CORP), 15 April 1998 (15.04.1998) -- --------	1-4

International patent classification (IPC)
H04R 1/28 (2006.01)
H04N 7/15 (2006.01)

Download your patent documents at www.prv.se
The cited patent documents can be downloaded at www.prv.se by following the links:

- In English/Searches and advisory services/Cited documents (service in English) or
- e-tjänster/anförda dokument(service in Swedish).

Use the application number as username.
The password is **NMJHUDMLLR**.

Paper copies can be ordered at a cost of 50 SEK per copy from PRV InterPat (telephone number 08-782 28 85).

Cited literature, if any, will be enclosed in paper form.

GB	2234137	A	23/01/1991	DE	3923740	C	06/12/1990
				FR	2650466	A,B	01/02/1991
				GB	9015256	D	00/00/0000
US	120030228025	A1	11/12/2003	US	D477930	S	05/08/2003
				US	7146014	B	05/12/2006
EP	0836324	A2	15/04/1998	CA	2218002	A,C	09/04/1998
				DE	69731864	D,T	03/11/2005
				JP	10285569	A	23/10/1998