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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 March 2008 (06.03.2008)

PCT

(10) International Publication Number
WO 2008/025059 A1

(51) International Patent Classification:
G06F 19/00 (2006.01)

(21) International Application Number:
PCT/AU2007/001214

(22) International Filing Date: 23 August 2007 (23.08.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2006904660 28 August 2006 (28.08.2006) AU
2007903280 19 June 2007 (19.06.2007) AU

(71) Applicant *(for all designated States except US)*: **E-NOSE PTY LTD** [AU/AU]; 145 National Innovation Centre, Australian Technology Park, Eveleigh, New South Wales 1430 (AU).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BELL, Graham** [AU/AU]; 6A Barons Crescent, Hunters Hill, New South Wales 2110 (AU). **HIBBERT, Brynn** [AU/AU]; 1/49 Church Street, Randwick, New South Wales 2031 (AU).

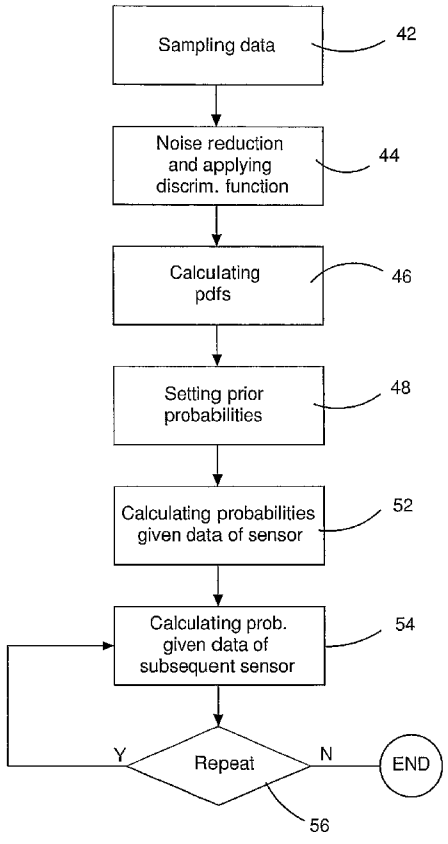
(74) Agent: **GRIFFITH HACK**; Level 19, 109 St Georges Terrace, Perth, Western Australia 6000 (AU).

(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,

[Continued on next page]

(54) Title: A METHOD OF DETERMINING THE PROBABILITY THAT DATA IS ASSOCIATED WITH A SOURCE OF A PLURALITY OF SOURCES



(57) Abstract: The present invention provides a method and apparatus for determining the probability that sampled data, associated with a source and obtained from a plurality of data input sources, are from a known source, given prior observed data obtained by the plurality of data input sources for one or more known sources. In one embodiment the data input sources are sensors for detecting molecules conveyed though the air and the method comprises identifying a source of the molecules. The present invention also provides a method of determining information about the position of at least one sensor relative to a previously known type of source.

WO 2008/025059 A1

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*

A METHOD OF DETERMINING THE PROBABILITY THAT DATA IS ASSOCIATED WITH A SOURCE OF A PLURALITY OF SOURCES

Field of the Invention

The present invention relates to a method of determining the probability that data is associated with a source of a plurality of sources, in particular to a method of predicting the source of data sampled with an array of sensors.

Background of the Invention

The Bayesian theory to discern different hypotheses when given experimental evidence (data) was developed by Rev. Thomas Bayes in 1763. Bayes' theorem allows calculation of the probability of an hypothesis H_i based on available evidence E. This is written $\Pr(H_i|E)$.

According to Bayes, for N competing and mutually exclusive hypotheses:

$$\Pr(H_i \mid E) = \frac{\Pr(E \mid H_i)\Pr(H_i)}{\sum_{j=1}^{N} \Pr(E \mid H_j)\Pr(H_j)} \quad (1)$$

where $\Pr(H_i|E)$ is the probability of the hypothesis H given evidence E, and $\Pr(H_i)$ is the prior probability (before any evidence) of the hypothesis and the sum is over all mutually exclusive hypotheses. The term $\Pr(E|H_i)$ is either the probability of finding the evidence E given the truth of the hypothesis H_i, or the probability density function

of the hypothesis H_i at a value given by the evidence E (this is also known as the 'likelihood'). In the absence of any other prior knowledge of the system, each $Pr(H_i)$ can be set as $1/N$, the so-called "flat prior".

The inventors of the present invention have applied Bayesian theory to develop a statistical tool for the treatment in real time of data from unknown sources which allows for comparison of the real time data with a template of previously recorded data from known sources in order to effect a decision with a given probability of the source of the real time data being generated by one of the known sources.

Summary of the Invention

The present invention provides in a first aspect a method of determining the probability that sampled data, associated with a source and obtained from a plurality of data input sources, are from a known source, given prior observed data obtained by the plurality of data input sources for one or more known sources, the method comprising the steps of:

a) calculating probability density functions for each known source for the sampled data being associated with that source and/or calculating probability density functions for the sampled data being associated with none of the known sources using the data and the prior observed data;

b) calculating a first set of probabilities that the sampled data is associated with at least one of the known sources or none of the known sources using a predetermined

prior probability and the probability density functions from step a); and

c) calculating a subsequent set of probabilities that the sampled data is associated with at least one of the known sources or none of the known sources using the probabilities calculated in the step b) and the probability density functions from step a) calculated from the sampled data.

In one embodiment of the invention, the method comprises repeating step c) for a further data set to calculate a further subsequent set of a plurality of probabilities.

In a further embodiment the method comprises iteratively repeating step c) to calculate further subsequent sets of pluralities of probabilities for a plurality of further data sets.

The further data set or the plurality of further data sets may be sampled prior, simultaneously or subsequently to the sampled data.

In one embodiment the data input source is a sensor. The plurality of sensors sample data from the source, each sensor being independent of the other sensors.

In an alternative embodiment the plurality of data input sources can correspond to data obtained from a spectrum, chromatogram or mass spectrum and so forth, optionally transformed to a single measure.

The method of the present invention typically relies on prior knowledge of a pattern of output data expected for

each known source for a given sensor or data input source. It is desirable, therefore that sufficient observed data should have been previously acquired to capture substantially all possible variations in data values measured by the given sensor for the respective known source, such that the observed data is sufficiently representative of the known source.

In one embodiment of the invention, the prior observed data has been mathematically manipulated to reduce noise, adjust the baseline, or smooth the prior observed data using suitable filter routines.

In an alternative embodiment, the prior observed data can be mathematically manipulated according to a discriminating function that is expected to discriminate between known sources of the data ($f(\mathbf{V})$). For example, the data values observed with an array of sensors V_1, V_2 ... V_{n-1} can be divided by the data values observed with sensor V_n. Other discriminating functions include, but are not limited to, logarithms of data values or performing a principal component analysis (PCA) and taking PC scores, or other linear or non-linear combinations of the data observed with the array of sensors. The number of discriminating functions may be less than or equal to the number of sensor outputs. For example, in the case where the data observed with the array of sensors V_1, V_2 ... V_{n-1} can be divided by the data observed with sensor V_n, there are one fewer discriminating functions than data outputs.

Typically, the data associated with the source will undergo the same mathematical manipulation as the prior observed data to reduce noise, adjust the baseline for the

sampled data, or smooth the data, and/or to obtain discriminating functions of the data as described above with respect to the prior observed data. For example, the data may be normalized.

In one embodiment of the invention, the prior observed data or the discriminating functions obtained from the prior observed data are described with a distribution function. The distribution function can be normal, lognormal, beta, and so forth.

For example, in one embodiment of the invention, the probability density functions (pdfs) for a given sensor for each known source are generated by the Microsoft Excel spreadsheet computer program with NORMDIST(x, μ, σ, 0), where x is the value of interest of a function. μ is the mean and σ is the standard deviation of the sensor data calculated using the spreadsheet functions AVERAGE(range) and STDEV(range). In alternative embodiments of the invention the pdfs for each known source may be generated by constructing a histogram of the data (a plot of frequency versus value for suitably defined ranges) or from the data by the kernel density method. An equivalent function may be used for different spreadsheet computer programs.

In one embodiment the pdfs for the sampled data for none of the known sources are typically generated as a square distribution outside the range of all the observed data for a particular function. It is given an integrated area of 0.9 over a range size equal to the total width of the known sources, with the remaining 0.1 inside the range of the observed data. These limits are set arbitrarily, but

the method of the present invention typically is insensitive to the exact values. The pdfs for the sampled data typically indicate that, if the value of a function during operation lies outside the range of all known sources, the probability that it belongs to none of the sources is 0.9. In this case there is a small (0.1) probability that even if a result is located in the range expected the source is still not from one of the known sources.

In a further embodiment of the invention, the probability density functions for each known source and the probability density functions that the sampled data is associated with none of the known sources are normalized.

In one embodiment of the invention the predetermined prior probability is a flat prior probability. In an alternative embodiment of the invention the predetermined prior probability is set according to existing information about the likely source of the data.

In one embodiment of the invention the step of calculating the plurality of respective probabilities that the data are associated with at least one of the known sources or none of the known sources, given the data obtained from a first of the data input sources or sensors, comprises applying Bayesian theory. For each source, the pdf calculated from data obtained by the first of the data input sources or sensors is multiplied with the predetermined prior probability, and the resulting respective products are divided over the sum of all the products for each source. It will be understood that that

the term "each source" refers to each and none of the known sources.

Alternatively, for each source, the pdf calculated from the discriminating function related to data sampled by the first of the data input sources or sensors is multiplied with the predetermined prior probability, and the resulting respective products are divided over the sum of all the products for each source.

In another embodiment of the invention the step of calculating a plurality of respective probabilities that the sampled data is associated with at least one of the known sources or none of the known sources, given the data sampled by the first of the data input sources or sensors and a subsequent data input source or sensor, also comprises applying Bayesian theory. For each source, the pdf calculated from data sampled by the subsequent data input source or sensor is multiplied with the respective probabilities calculated in the previous step, and the resulting respective products are divided over the sum of all the products for each source.

Alternatively, for each source, the pdf calculated from the discriminating function related to subsequently sampled data is multiplied with the respective probabilities calculated in the previous step, and the resulting respective products are divided over the sum of all the products for each source.

In one embodiment of the invention the method can be adapted for recognition or identification of the source of the sampled data where the calculated probability that the

data sampled with the first and subsequent data input sources or sensors is associated with a known source is equal to or above a threshold probability.

In a second aspect of the invention there is provided a system of determining the probability that data, associated with a source and sampled by a plurality of data input sources, is from a known source, given prior observed data measured by the plurality of data input sources for one or more known sources, the system comprising:

a) a first calculator for calculating probability density functions for each known source for the sampled data being associated with that source and/or calculating probability density functions for the sampled data not being associated with any one of the known sources using the sampled data and the prior observed data;

b) a second calculator for calculating a first set of probabilities that the sampled data is associated with at least one of the known sources or none of the known sources using a predetermined prior probability and the probability density functions from step a); and

c) a third calculator for calculating a subsequent set of probabilities that the sampled data is associated with at least one of the known sources or none of the known sources using the probabilities calculated in step b) and the probability density functions from step a) calculated from the sampled data.

The present invention provides in a third aspect a computer program arranged for determining the probability that data associated with a source and sampled by a

plurality of data input sources is from a known source in accordance with the first aspect of the present invention.

In a further aspect of the invention there is provided a computer readable storage medium comprising the computer program as defined above.

In the first, second and third aspects of the present invention the data typically is sampled by detecting molecules conveyed through a gaseous medium emitted from one or more sources. For example, the molecules may be associated with an odour.

In a fourth aspect of the invention there is provided a method of identifying the source of a predetermined type of molecule conveyed through a gaseous medium and sampled by a plurality of sensors, given prior observed data measured by the plurality of sensors in respect of the predetermined type of molecule from one or more known sources, the method comprising the steps of:

a) calculating probability density functions for each known source that the detected molecules are associated with that source and/or calculating probability density functions for the data being associated with none of the known sources using the sampled data associated with the detected molecules and the prior observed data;

b) calculating a first set of probabilities that the detected molecules originate from at least one of the known sources or none of the known sources using a predetermined prior probability and the probability density functions from step a);

c) calculating a subsequent set of probabilities that the detected molecules originate from at least one of the known sources using the probabilities calculated in step b) and the probability density functions from step a); and

d) assigning a source identity to the sampled data when one of the probabilities calculated in step c) previous step is equal to or above a threshold probability.

The step of assigning the source identity typically comprises assigning the source identity to the sampled data that has the greatest probability.

The present invention provides in a fifth aspect a system for identifying the source of a predetermined type of molecule conveyed through a gaseous medium and sampled by a plurality of sensors, given prior observed data measured by the sensors in respect of the predetermined type of molecule from one or more known sources, comprising:

a plurality of sensors for detecting the molecules conveyed through the gaseous medium, each sensor being arranged for detecting the molecules of the predetermined type to detect a quantity of the molecules and to produce electronic data associated with the quantity of the molecules; and

a processor for:

a) calculating probability density functions for each known source that the detected molecules originate from that source and/or probability density functions for detected molecules not being associated with any one of the known sources using the sampled data associated with the detected molecules and the prior observed data;

b) calculating a first set of probabilities that the detected molecules originate from at least one of the known sources or none of the known sources using a predetermined prior probability and the probability density functions from step a);

c) calculating a subsequent set of probabilities that the detected molecules originate from at least one of the known sources or none of the known sources using the probabilities calculated in step b) and the probability density functions from step a); and

d) assigning a source identity to the detected molecules when one of the probabilities calculated in step c) is equal to or above a threshold probability.

The processor of the system typically is arranged so that the step of assigning a source identity comprises assigning the source identity to the sampled data that has the greatest probability.

The present invention provides in a sixth aspect a method of determining information about the position of at least one sensor relative to a previously known type of source with which data is associated, the data being relating to a property that is detectable at a plurality of locations and the property being dependent on a distance from the at least one source, the method comprising the steps of:

obtaining data at the plurality of locations using the at least one sensor;

determining the probability that the obtained data is associated with a previously known type of source, the

probability being calculated for each of the plurality of locations;

selecting, dependent on the calculated probabilities, at least one previously known type of source; and

comparing the data for the plurality of locations and from relative differences between the data determining information about a position of the at least one sensor relative to the location of the at least one selected source.

For example, the property may be associated with molecules conveyed through a gaseous medium such as those that relate to an odour that may originate from a sewage plant, a meat processing plant, a chemical plant or the like.

In one embodiment the step of comparing the data comprises determining at least the approximate location of the selected source of the data. For example, the type of source may be known and/or the source may be known, but the location of the source may be unknown. The selected source may be a moving source or may have moved to an unknown location. For example, the source may be positioned on a ship or on another moving object and the source may have emitted molecules that are conveyed through the air and that signal an emergency. The apparatus may then be used to locate the ship or the like.

The step of determining the probabilities that the obtained data is of a previously known type typically comprises use of the method in accordance with the first aspect of the present invention.

The step of comparing the data typically comprises comparing data intensities.

In one embodiment of the present invention the at least one sensor is positioned at each of the plurality of locations.

In another embodiment of the present invention, the at least one sensor moves from one location to another and the data is detected successively. In this case the method typically also comprises, in a feedback loop, estimating a direction or region of the selected source. This step typically is conducted so that the information provided by the feedback loop controls the movement of the at least one sensor so that the at least one sensor moves towards the selected source or away from the selected source.

Moving the at least one sensor may be effected by shifting, driving, flying or otherwise moving the at least one detection system.

The present invention comprises in a seventh aspect an apparatus for determining information about the position of at least one sensor relative to a previously known type of source with which data is associated, the data relating to a property that is detectable at a plurality of locations and the property being dependent on a distance from the at least one source, the apparatus comprising:

at least one sensor for obtaining the data at the plurality of locations; and

a system for:

determining the probability that the obtained data is associated with a previously known type of source, the probability being calculated for each of the plurality of locations;

selecting, dependent on the calculated probabilities, at least one of the previously known type of source; and

comparing the data for the plurality of locations and from relative differences between the data determining information about a position of the at least one sensor relative to the location of the at least one selected source.

The system typically comprises the system for determining the probabilities that the obtained data is associated with a previously known source in accordance with the second aspect of the present invention. Further, the system typically is arranged for determining at least the approximate location of the selected source.

For example, the property may be associated with molecules conveyed through a gaseous medium such as those that relate to an odour that may originate from a sewage plant, a meat processing plant, a chemical plant or the like. Each sensor may be a sensor array.

The apparatus may comprise a plurality of sensors arranged for positioning at respective positions.

In another embodiment of the present invention, the at least one sensor is arranged to move from one location to another and the data is detected successively.

The at least one sensor may be arranged for driving, flying or otherwise moving. The at least one sensor may move separately from the system or together with the system. The apparatus may also comprise a transmitter for transmitting information to a base and a receiver for receiving information or a control signal from the base, which may be used for influence of the movement, the data detection or any other property.

The apparatus may be arranged so that the movement is controlled via the feedback loop which is arranged so that the at least one sensor moves towards the selected source.

Alternatively, the apparatus may be arranged so that the movement is controlled via the feedback loop so that the at least one sensor moves away from the selected source.

The system typically comprises a computer and computer software for operation of the method according to the first aspect of the present invention.

The invention will be more fully understood from the following description of specific embodiments of the invention. The description is provided with reference to the accompanying drawings.

Brief Description of the Drawings

Figure 1 shows a schematic representation of a system for detecting a predetermined type of molecule conveyed through a gaseous medium;

Figure 2 is a graph of mathematically manipulated representative observed data from four known sources, and

sampled by an array of five sensors, showing the average ratios of output data from sensors 1, 2, 3, and 5, respectively to the output data from sensor 4 and standard deviations thereof by location and sensor;

Figure 3 shows a schematic block diagram illustrating a system according to one embodiment of the invention for determining the probability that data from an unknown source and sampled by an array of sensors is from a known source, given prior observed data measured by the array of sensors from one or more known sources;

Figure 4 shows a flow chart illustrating the steps of a method of determining the probability that data from an unknown source and sampled by an array of sensors is from a known source, given prior observed data measured by the array of sensors from one or more known sources, according to one embodiment of the invention;

Figure 5 shows a spreadsheet of representative sampled and prior observed data from four known sources, and sampled by an array of sensors, wherein the sampled and prior observed data have been treated according to a specific embodiment of the method of the present invention; and

Figure 6 shows a schematic representation illustrating an apparatus and a method according to an embodiment of the present invention.

Detailed Description of Specific Embodiments

A method of determining the probability that data from an unknown source, and sampled by an array of sensors, is from a known source, given prior observed data measured by the array of sensors from one or more known sources, is

now described with reference to the following illustrative example.

In this specific example the sampled and prior observed data relates to a voltage output from an array of electronic detectors configured to detect predetermined types of molecules conveyed through a gaseous medium. Typically the molecules of the predetermined type result in an airborne odour and comprise organic molecules.

Figure 1 shows a schematic representation of a system for detecting a predetermined type of molecule conveyed through a gaseous medium 10. The system 10 comprises one or more detectors 12 and a processor 14. In this embodiment there are four detectors 12 each of which is positioned in proximity of a respective known odour sources 16a, 16b, 16c, and 16d of a plant 16. In a preferred embodiment the processor 14 is a microprocessor or a computer.

The plant can be any source that generates organic molecules entering the environment, thereby causing an odour. For example, the plant may be a sewage plant, a meat producing or livestock processing plant, brewery, chemical plant, oil refinery and so forth. In this embodiment the detectors 12 are multi channel array detectors that comprise an array of five chemical sensors. Each sensor is an odour sensor that has a surface which comprises a receptor for the predetermined type of molecule. When a particular type of odour molecule interacts with a receptor molecule, an electrical property of the sensor surface changes and therefore the odour molecule can be detected. Such multi channel array

detectors are also known as an "e-nose" or "electronic nose". Each detector also has a separate detector channel for detecting a temperature and a separate detection channel for detecting humidity.

For example, the detectors 12 may comprise surface acoustic wave sensors, conducting polymer sensors, quartz crystal microbalance sensors, fibre optic sensors, metal oxide sensors.

The detectors 12 are arranged to produce electronic data that is associated with the airborne odour. In this illustrative example the detectors 12, positioned in proximity of known odour sources 16a, 16b, 16c, 16d, are exposed to the airborne odour from the plant 16. The detectors 12 then produce electronic data that is associated with a quantity and type of airborne odour specific to the known sources 16a, 16b, 16c, 16d. The electronic data is directed to processor 14 which is, positioned remote from the detectors 12.

The processor 14 processes the electronic data to provide information about the airborne odour. Additionally, the processor 14 may be adapted to store, accumulate and update information about the airborne odours from respective known sources 16a, 16b, 16c, 16d.

Typically, sensor measurements are taken in predetermined intervals such as every 0.5 seconds and then sequentially converted into electronic data and sent to the processor 14. Prior observed data relating to known sources 16a, 16b, 16c, 16d may be recorded over an extended period (hours, days or weeks) explore natural variation of the

odour from that source to be represented in the data. Alternatively, the data may be obtained rapidly. For example the moving device could be programmed to record a strong and/or consistently intense odour, and the processor 14 could then provide the probability of the associated data corresponding to any one of the known sources. If this is not the case, the processor 14 could then be programmed to include that data in a data library and move towards the corresponding source (or away from the corresponding source). The corresponding source that has been located this way could then be included as a known source.

The voltage output from the sensors is averaged to smooth noise. Further, if the sensors are not zeroed (i.e. calibrated to give 0 V in the absence of an odour), it is desirable to subtract the baseline (i.e. the data obtained when there is no odour) from all readings.

Additionally, the voltage output is mathematically manipulated with a discriminating function $f(x)$ to better discriminate among the known odour sources 16a, 16b, 16c, 16d. In this specific example the sensor outputs of sensors 1, 2, 3, and 5 have been divided by the sensor output of sensor 4. A graph of the average ratios and standard deviations by location and sensor of the prior observed data in this specific example is shown in Figure 2.

Referring now to Figures 3, 4 and 5, a specific example to illustrate an application of the method of the invention is now described.

Figure 3 shows a schematic block diagram of a system for identifying the source of a predetermined type of molecule conveyed through a gaseous medium 20. The system 20 includes the detector 12 and the processor 14 described with reference to Figure 1.

The processor 14 includes a first calculator 22, a second calculator 24, a third calculator 26, and an identity assignor 28.

The first calculator 22 calculates probability density functions for each known source and probability density functions for none of the known sources using the data sampled by the sensors of detector 12 and the prior observed data. It is to be understood that prior to calculating the probability density functions for each source, the first calculator 22 may initially average the sampled data to smooth noise and mathematically manipulate the smoothed data with function $f(x)$. For example, the sensor outputs of sensors 1, 2, 3, and 5 can be divided by the sensor output of sensor 4.

The second calculator 24 calculates a first set of a plurality of respective probabilities that the predetermined type of molecule originates from each one and none of the known sources using a predetermined prior probability and the probability density functions calculated by the first calculator 24 from data sampled by the first of the sensors.

The third calculator 26 calculates a subsequent set of a plurality of respective probabilities that the predetermined type of molecule originates from each one

and none of the known sources using the respective probabilities calculated by the second calculator and the probability density functions calculated by the first calculator 24 from data sampled by the subsequent sensor.

It will be understood that the system can include one or more subsequent calculators to calculate a further subsequent set of a plurality of respective probabilities that the predetermined type of molecule originates from each one and none of the known sources using the respective probabilities calculated by the preceding calculator and the probability density functions calculated by the first calculator 24 from data sampled by the further subsequent sensor(s).

The identity assignor assigns a known source identity to the sampled data when one of the respective probabilities calculated by the third calculator 26, or one of the subsequent calculators, is equal to or above a threshold probability.

Figure 4 shows a flow chart 40 illustrating the steps of a method of determining the probability that data from an unknown source and sampled by an array of sensors is from a known source, given prior observed data measured by the array of sensors from one or more known sources, and Figure 5 shows a spreadsheet 60 of representative sampled and prior observed data. In the spreadsheet 60 known source 16a is labeled S, known source 16b is labeled B, known source 16c is labeled D, and known source 16d is labeled sb.

Data representing an odour from an unknown source in plant 16, is sampled by sensors 1-5 of the detector 12 (step 42). In step 44 the sampled data is initially averaged to smooth noise and mathematically manipulated with discriminating function $f(x)$ (e.g. the sensor outputs of sensors 1, 2, 3, and 5 have been divided by the sensor output of sensor 4) in the first calculator 22. A representative sample of sampled data ratios 62 is shown for purposes of illustration only in the spreadsheet of Figure 5.

The sampled data ratios 62 relating to sensors 1, 2, 3, and 5 are then converted in the first calculator 22 to respective probability density functions for each known source 16a, 16b, 16c, 16d and probability density functions for none of the known sources using the prior observed data, as shown in Figure 2, with NORMDIST in Excel (step 46). A representative sample of respective probability density functions for each known source 16a, 16b, 16c, 16d and probability density functions for none of the known sources 64 is shown for purposes of illustration only in the spreadsheet of Figure 5.

In the absence of any better information the predetermined prior probabilities 68 are set as $1/N$, the so-called "flat prior" (step 48). In this specific example the predetermined prior probabilities 68 are 0.2 for each source.

The respective pdfs for each source for data relating to sensor 1 (64a) are then multiplied with the predetermined prior probabilities 68, in this specific example 0.2, and the resulting respective products are then divided by a

product of the predetermined prior probability 68 and the sum of the pdfs from each source for data relating to sensor 1 (step 52).

A mathematical representation of the calculation performed by the second calculator 24 in step 52 is shown below:

$$\Pr(16a \mid x_i) = \frac{\Pr(x_i \mid 16a)\Pr(16a)}{[\Pr(x_i \mid 16a) + \Pr(x_i \mid 16b) + \Pr(x_i \mid 16c) + \Pr(x_i \mid 16d) + \Pr(x_i \mid none)]\Pr(16a)}$$

where $\Pr(x_1|16a)$ is the pdf from step 46 at function value x_1 (e.g. x_1 is sensor 1/sensor 4 ratio of 1.479) for known source 16a, and so forth. The result $\Pr(16a|x_1)$ is read as "the probability of the odour coming from known source 16a, given the function value x_1". Pr(16a) = predetermined prior probability = 0.2.

In step 54 of the method of the present invention, the respective pdfs 64b for each source for data relating to sensor 2 are then multiplied with the respective results 70a of the previous step for each source, in other words $\Pr(16a|x_1)$, $\Pr(16b|x_1)$, $\Pr(16c|x_1)$, $\Pr(16d|x_1)$, and $\Pr(none|x_1)$, replacing the predetermined prior probability 68. The resulting respective products are then divided by a sum of the products of 70a with the respective pdfs 64b from each source for data relating to sensor 2. The results of this calculation in the third calculator 26 are $\Pr(16a|x_1,x_2)$, $\Pr(16b|x_1,x_2)$, $\Pr(16c|x_1,x_2)$, $\Pr(16d|x_1,x_2)$, and $\Pr(none|x_1,x_2)$, which are read as "the respective probabilities of the odour coming from known sources 16a, 16b, 16c, 16d, and none of the known sources given the function value x_1 *and the* function value x_2" (70b).

At this stage one can make a decision 56 to cease further calculation, in particular if the probabilities relating to any one of each source has fallen below a low threshold probability so that the source in question can be disregarded as the true source of the odour, or if the probabilities relating to any one of each source is equal to or has risen above a high threshold probability so that the source having the greatest probability can be identified as the true source of the odour.

Conversely, one can continue to iteratively repeat step 54 with the respective pdfs 64c, 64d for each source relating to data obtained by the subsequent sensors 3 and 5 to obtain the respective probabilities of the odour coming from known sources 16a, 16b, 16c, 16d, and none of the known sources given the data from sensors 1, 2 and 3 (70c), and the respective probabilities of the odour coming from known sources 16a, 16b, 16c, 16d, and none of the known sources given the data from sensors 1, 2, 3 and 5 (70d).

It is evident from 70d in the spreadsheet in Figure 5 that the iterative process of the present invention has significantly discriminated between the available hypotheses for the source of the odour. The probabilities 70d calculated by iteratively repeating step 54 indicate a 90% probability that the odour's source was known source 16b, and with respect to the alternative possible sources, at most only a 6% probability that the odour's source was known source 16a.

Referring now to Figure 6, an apparatus and a method according to a specific embodiment of the present

invention are now described. The apparatus 80 comprises in this embodiment the system 10 for detecting a predetermined type of molecule conveyed through a gaseous medium.

Further, the apparatus 80 comprises the system for determining the probability that data associated with a source and obtained from a plurality of data input sources is associated with a known source. The system is analogous to that described above with reference to Figures 3-5.

The apparatus 80 is arranged to obtain data using the system 10. The obtained data is then processed to determine the probabilities that the obtained data is from a previously known source. In this embodiment the apparatus 80 is arranged to detect an odour. The odour may be emitted from a plurality of sources such as source 82. The apparatus 80 is in this embodiment movable relative to the odour sources. Areas 84 and 86 indicate locations of the apparatus 80. When the apparatus 80 is at position 84, data is obtained and the probability for the odour to originate from an odour sources 82 or 88 is calculated. The calculation may have resulted in a high probability for the odour to originate from source 82, in which case the apparatus is set-up to detect changes in intensity associate with that odour. When the apparatus 80 moves to position 86, that intensity reduced as position 86 is located slightly further distant from the odour source 82. The detected change in odour intensities are compared and the apparatus 80 then detects that the position 86 is further away from the source of the order 82 than the position 84.

The apparatus 80 is in this embodiment arranged flying towards the source 82. As the apparatus 80 has detected that the position 86 is further away from the source 82 then the position 84, the apparatus 80 generates a flight path control signal which effects movement of the apparatus 80 towards the source 82 to position 90. At position 90 further data is obtained and the odour intensities are compared and, in a feedback loop, the obtained data is used to control the flight path of the apparatus 80 towards the source 82 whereby the location of the source 82 is identified.

It is to be appreciated that in variations of the described embodiment the apparatus 80 may also be arranged for driving, floating or for any other type of movement. Further, the apparatus 80 may be one of a plurality of apparatus of the same type which are located in the proximity of odour sources and which may not be movable. In this case data obtained by each of the plurality of apparatus is processed and used to identify the location of the odour source 82.

Further, it is to be appreciated that the apparatus 80 may alternatively be directed so that the apparatus 80 moves away from a source of odour. In this case the movement of the apparatus 80 is guided by a decrease in intensity of the data associated with the odour originating from an identified source.

Even though the invention has been described with reference to a particular example it will be appreciated by those skilled in the art that the invention may be

embodied in many other forms. For example, the system 10 may comprise any number of detectors 12. Further, each detector 12 may be any type of detector with any number of sensors sensitive to data of any particular type generated by any source.

In the claims which follow and in the preceding description of the invention, except where the context requires otherwise due to express language or necessary implication, the word "comprise" or variations such as "comprises" or "comprising" is used in an inclusive sense, i.e. to specify the presence of the stated features but not to preclude the presence or addition of further features in various embodiments of the invention.

THE CLAIMS:

1. A method of determining the probability that sampled data, associated with a source and obtained from a plurality of data input sources, are from a known source, given prior observed data obtained by the plurality of data input sources for one or more known sources, the method comprising the steps of:

a) calculating probability density functions for each known source for the sampled data being associated with that source and/or calculating probability density functions for the sampled data being associated with none of the known sources using the data and the prior observed data;

b) calculating a first set of probabilities that the sampled data is associated with at least one of the known sources or none of the known sources using a predetermined prior probability and the probability density functions from step a); and

c) calculating a subsequent set of probabilities that the data is associated with at least one of the known sources or none of the known sources using the probabilities calculated in the step b) and the probability density functions from step a) calculated from the sampled data.

2. The method according to claim 1, wherein the method comprises repeating step c) for a further data set to calculate a further subsequent set of a plurality of probabilities, the further data set having been sampled before, simultaneously or after the sampled data.

3. The method according to claim 1 or 2, wherein the method comprises iteratively repeating step c) to calculate further sets of pluralities of probabilities for each of a plurality of further data set obtained before, simultaneously or after the sampled data.

4. The method according to any one of the preceding claims, wherein the prior observed data has been mathematically manipulated to reduce noise or adjust the baseline.

5. The method according to any one of the preceding claims, wherein the prior observed data is mathematically manipulated according to a discriminating function that is expected to discriminate between known sources of the data ($f(\mathbf{V})$).

6. The method according to claim 5, wherein the discrimination functions include ratios of data values, logarithms of data values or ratios of data values, performing a principal component analysis (PCA) and taking PC scores, or other linear or non-linear combinations of the data observed with the array of sensors.

7. The method according to claims 5 or 6, wherein the data sampled from the source will undergo the same mathematical manipulation as the prior observed data to reduce noise, adjust the baseline for the sampled data, or remove or minimize randomness in the data, and/or to obtain discriminating functions of the data.

8. The method according to claims 5 or 6, wherein the prior observed data or the discriminating functions

obtained from the prior observed data are described with a distribution function.

9. The method according to claim 8, wherein the distribution function is normal, lognormal, beta.

10. The method according to any one of the preceding claims, wherein the predetermined prior probability is a flat prior probability.

11. The method according to any one of the preceding claims, wherein the method can be adapted for recognition or identification of the source of the data whereby the calculated probability that the data obtained from the first and subsequent data input sources originates from a known source is equal to or above a threshold probability.

12. A system of determining the probability that data, associated with a source and sampled by a plurality of data input sources, is from a known source, given prior observed data measured by the plurality of data input sources for one or more known sources, the system comprising:

a) a first calculator for calculating probability density functions for each known source for the sampled data being associated with that source and/or calculating probability density functions for the sampled data not being associated with any one of the known sources using the sampled data and the prior observed data;

b) a second calculator for calculating a first set of probabilities that the sampled data is associated with at least one of the known sources or none of the known

sources using a predetermined prior probability and the probability density functions from step a); and

c) a third calculator for calculating a subsequent set of probabilities that the sampled data is associated with at least one of the known sources or none of the known sources using the probabilities calculated in step b) and the probability density functions from step a) calculated from the sampled data.

13. A computer program arranged for determining the probability that data associated with a source and sampled by a plurality of data input sources is from a known source in accordance any one of claims 1 to 11.

14. A computer readable storage medium comprising the computer program as defined in claim 13.

15. A method of identifying the source of a predetermined type of molecule conveyed through a gaseous medium and sampled by a plurality of sensors, given prior observed data measured by the plurality of sensors in respect of the predetermined type of molecule from one or more known sources, the method comprising the steps of:

a) calculating probability density functions for each known source that the detected molecules are associated with that source and/or calculating probability density functions for the data being associated with none of the known sources using sampled data associated with the detected molecules and the prior observed data;

b) calculating a first set of probabilities that the detected molecules originate from at least one of the known sources or none of the known sources using a

predetermined prior probability and the probability density functions from step a);

c) calculating a subsequent set of probabilities that the detected molecules originate from at least one of the known sources using the probabilities calculated in step b) and the probability density functions from step a); and

d) assigning a source identity to the sampled data when one of the probabilities calculated in step c) previous step is equal to or above a threshold probability.

16. The method of claim 15 wherein the step of assigning the source identity comprises assigning the source identity to the sampled data that has the greatest probability.

17. A system for identifying the source of a predetermined type of molecule conveyed through a gaseous medium and sampled by a plurality of sensors, given prior observed data measured by the sensors in respect of the predetermined type of molecule from one or more known sources, comprising:

a plurality of sensors for detecting the molecule conveyed through the gaseous medium, each sensor being arranged for detecting the molecules of the predetermined type to detect a quantity of the molecules and to produce electronic data associated with the quantity of the molecules; and

a processor for:

a) calculating probability density functions for each known source that the detected molecules originate from that source and/or probability density functions for detected molecules not being

associated with any one of the known sources using the sampled data associated with the detected molecules and the prior observed data;

b) calculating a first set of probabilities that the detected molecules originate from at least one of the known sources or none of the known sources using a predetermined prior probability and the probability density functions from step a);

c) calculating a subsequent set of probabilities that the detected molecules originate from at least one of the known sources or none of the known sources using the probabilities calculated in step b) and the probability density functions from step a); and

d) assigning a source identity to the detected molecules when one of the probabilities calculated in step c) is equal to or above a threshold probability.

18. The system of claim 17 wherein the processor is arranged so that the step of assigning a source identity comprises assigning the source identity to the sampled data that has the greatest probability.

19. A method of determining information about the position of at least one sensor relative to a previously known type of source with data is associated, the data relating to a property that is detectable at a plurality of locations and the property being dependent on a distance from the at least one source, the method comprising the steps of:

obtaining data at the plurality of locations using the at least one sensor;

determining the probability that the obtained data is associated with the a previously known type of source, the probability being calculated for each of the plurality of locations;

selecting, dependent on the calculated probabilities, at least one previously known type of source; and

comparing the data for the plurality of locations and from relative differences between the data determining information about a position of the at least one sensor relative to the location of the at least one selected source of the data.

20. The method of claim 19 wherein the step of comparing the data comprises determining at least the approximate location of the selected source of the data.

21. The method of claim 19 or 20 wherein the step of determining the probabilities comprises using the method in accordance with the any one of claims 1 - 11.

22. The method of any one of claims 19 to 21 wherein the step of comparing the data comprises comparing data intensities.

23. The method of any one of claims 19 to 22 wherein the property is associated with molecules conveyed through a gaseous medium.

24. The method of any one of claims 19 to 23 wherein the at least one sensor is positioned at each of the plurality of locations.

25. The method of claim any one of claims 19 to 23 wherein the at least one sensor moves from one location to another and the data is detected successively.

26. The method of claim comprising, in a feedback loop, estimating a direction or region of the selected source.

27. The method of claim 26 conducted so that information provided by the feedback loop controls the movement of the at least one sensor so that the at least one sensor moves towards the selected source or away from the selected source.

28. The method of any one of claims 25 to 27 wherein moving comprises shifting, driving or flying.

29. An apparatus for determining information about the position of at least one sensor relative to a previously known type of source with which data is associated, the data relating to a property that is detectable at a plurality of locations and the property being dependent on a distance from the at least one source, the apparatus comprising:

at least one sensor for obtaining data at the plurality of locations; and

a system for:

determining the probability that the obtained data is associated with a previously known type of source, the probability being calculated for each of the plurality of locations;

selecting, dependent on the calculated probabilities, at least one previously known type of source; and

comparing the data for the plurality of locations and from relative differences between the data determining information about a position of the at least one sensor relative to the location of the at least one selected source of the data.

29. The apparatus of claim 28 the system comprises the system for determining the probabilities in accordance with claim 12.

30. The apparatus of claim 28 or 29 wherein the property is associated with molecules conveyed through a gaseous medium.

31. The apparatus of any one of claims 27 to 30 wherein the at least one sensor is a sensor array.

32. The apparatus of any one of claims 28 to 31 comprising a plurality of sensors arranged for positioning at respective positions.

33. The apparatus of any one of claims 28 to 32 wherein the at least one sensor is arranged to move from one location to another and the data is detected successively.

34. The apparatus of claim 33 wherein the apparatus is arranged so that the movement is controlled via a feedback loop which is arranged so that the at least one sensor moves towards, or away from, the selected source.

35. The apparatus of claim 33 or 34 wherein the apparatus is arranged so that the at least one sensor is driven or flown.

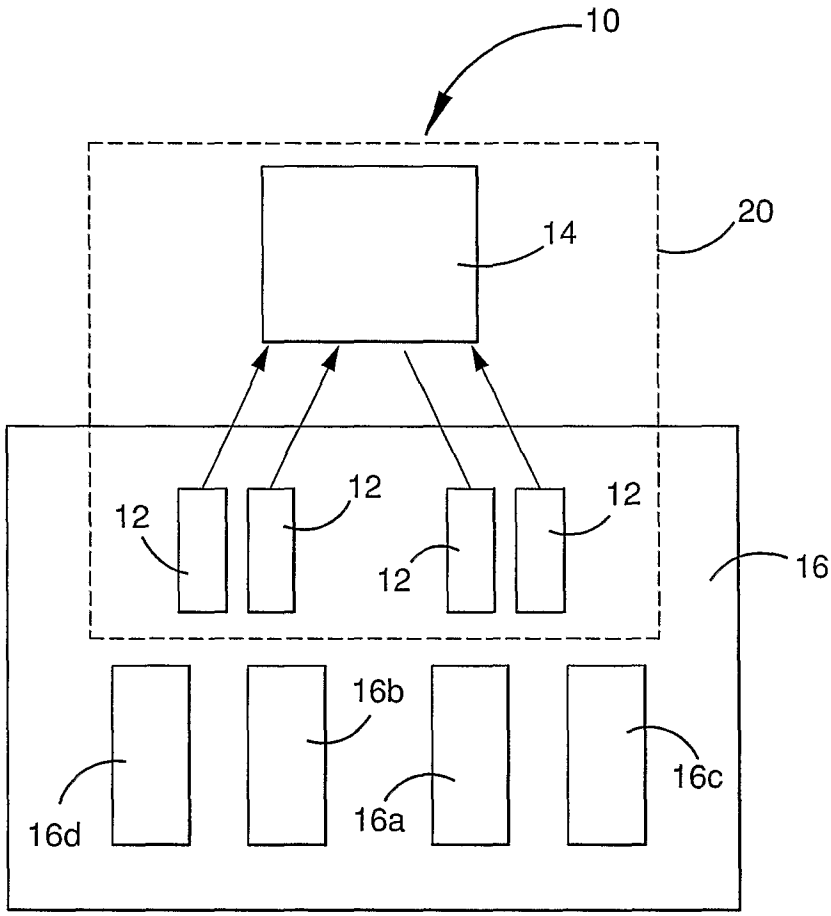


FIGURE 1

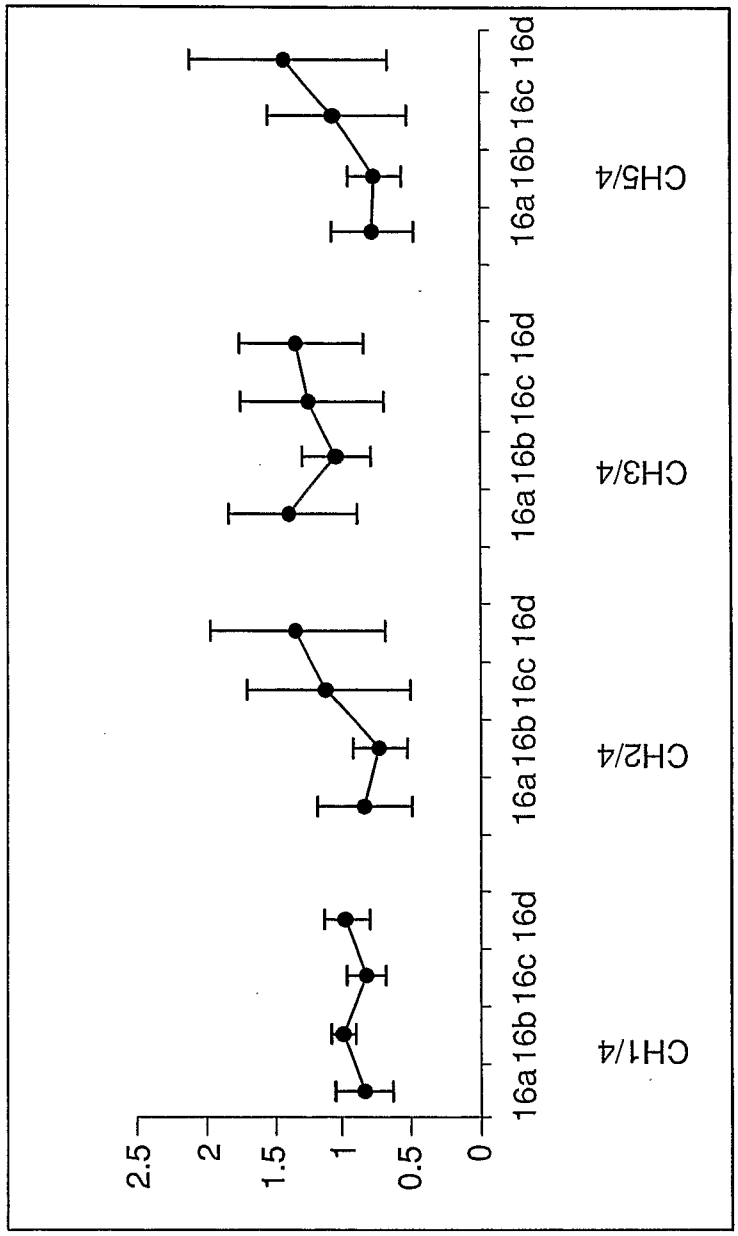

2.5
2
1.5
1
0.5
0
16a 16b 16c 16d
CH1/4
16a 16b 16c 16d
CH2/4
16a 16b 16c 16d
CH3/4
16a 16b 16c 16d
CH5/4

FIGURE 2

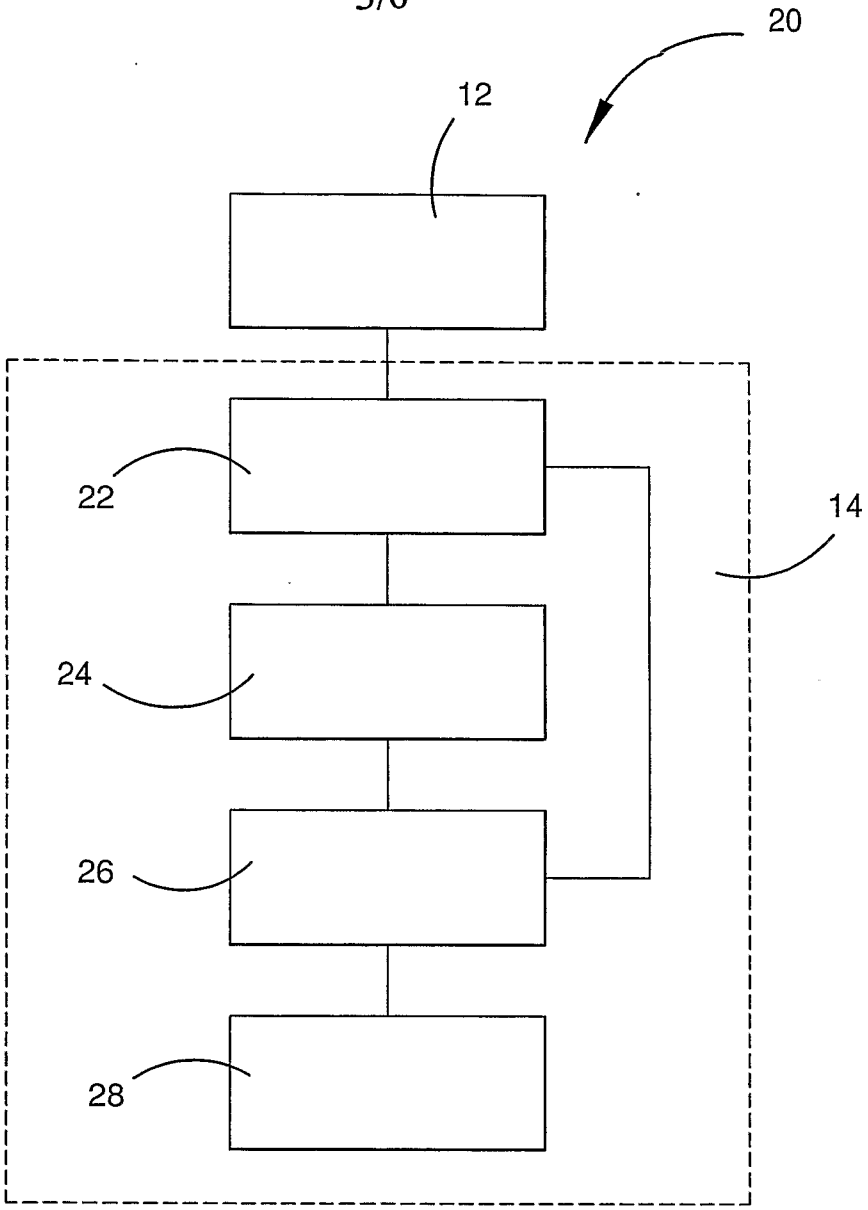


FIGURE 3

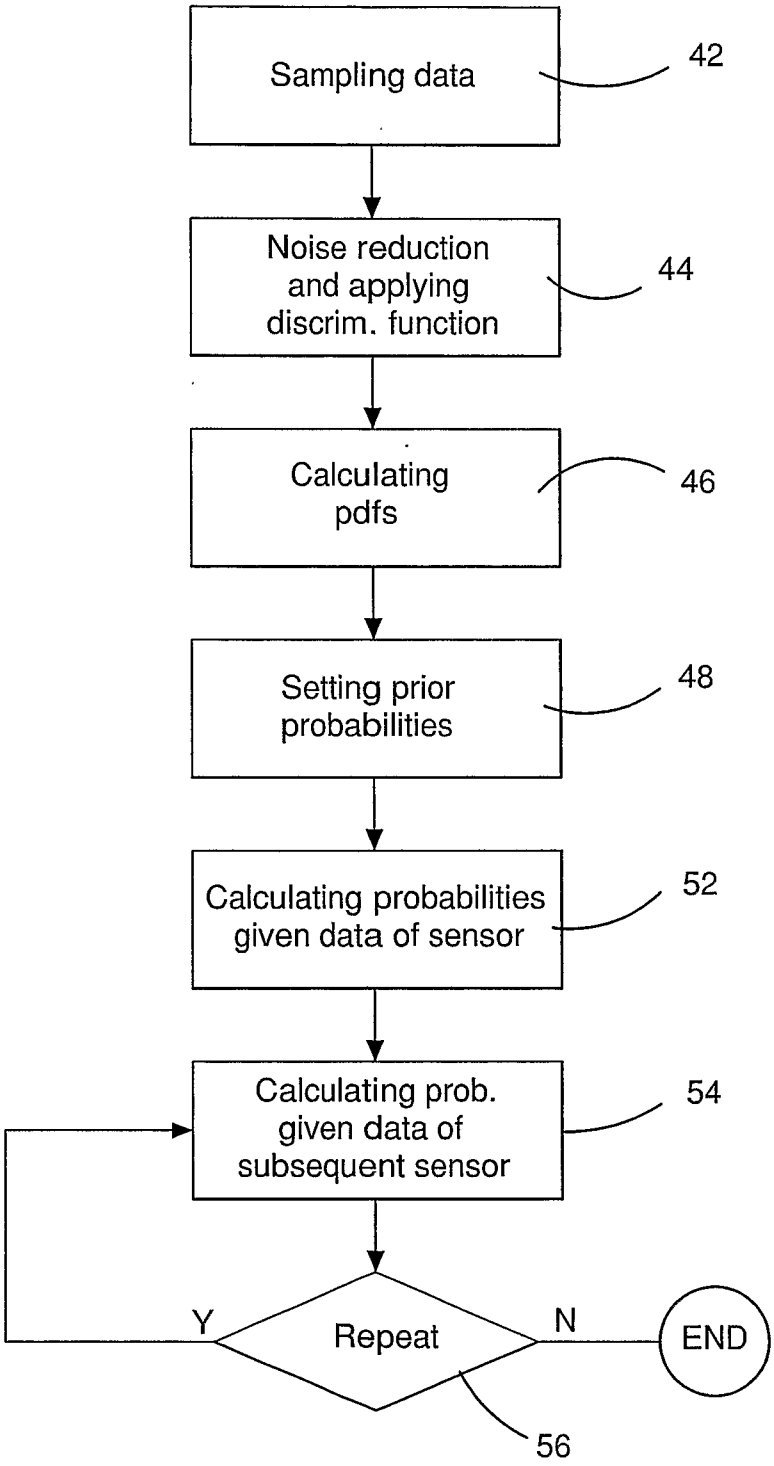

Sampling data
42
Noise reduction and applying discrim. function
44
Calculating pdfs
46
Setting prior probabilities
48
Calculating probabilities given data of sensor
52
Calculating prob. given data of subsequent sensor
54
Repeat
Y
N
END
56

FIGURE 4

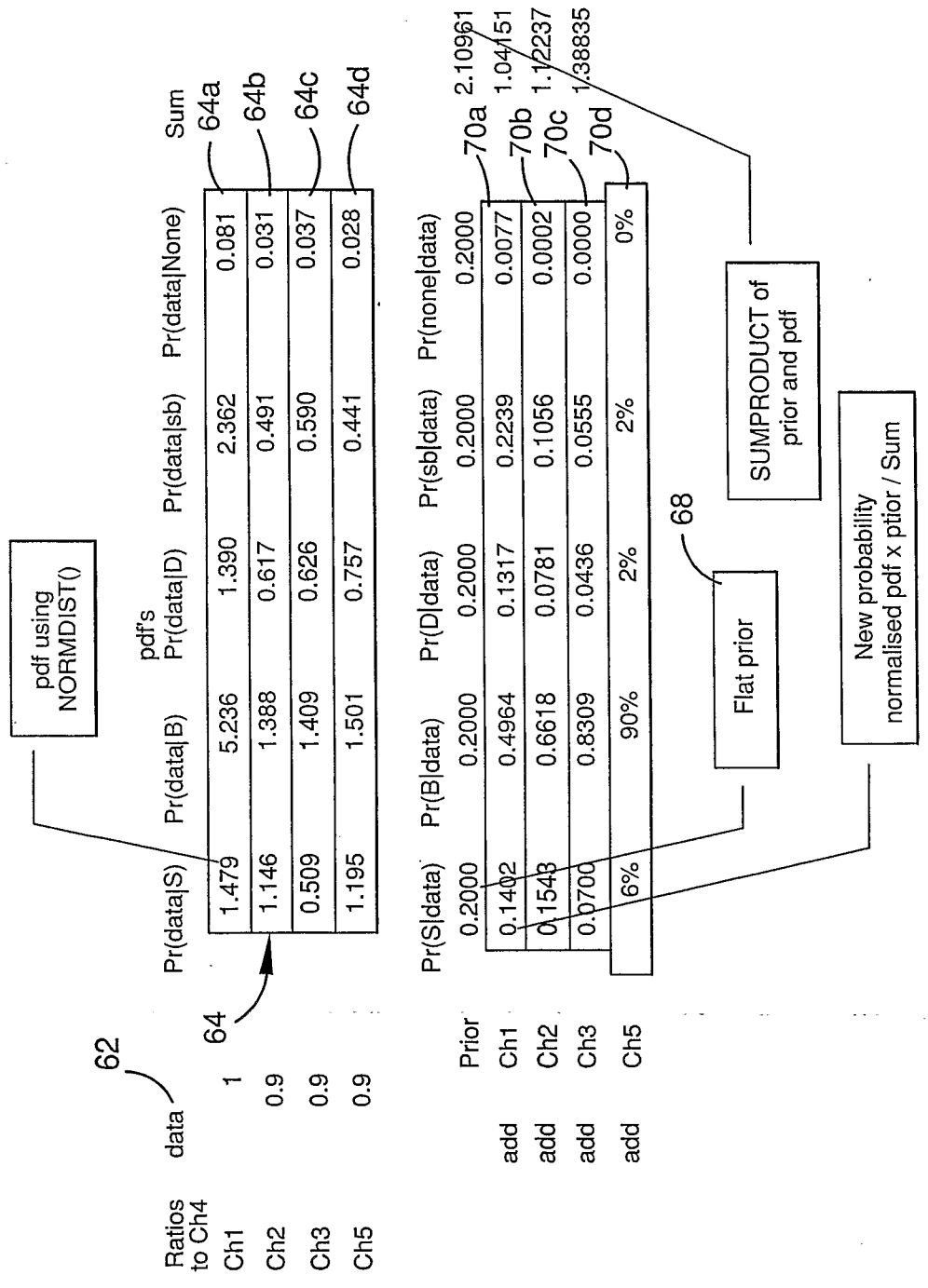

62
64
data
Ratios to Ch4
Ch1
Ch2
Ch3
Ch5
1
0.9
0.9
0.9
pdf using NORMDIST()
pdf's
Pr(data|S)
Pr(data|B)
Pr(data|D)
Pr(data|sb)
Pr(data|None)
Sum
1.479
5.236
1.390
2.362
0.081
64a
1.146
1.388
0.617
0.491
0.031
64b
0.509
1.409
0.626
0.590
0.037
64c
1.195
1.501
0.757
0.441
0.028
64d
Prior
Pr(S|data)
Pr(B|data)
Pr(D|data)
Pr(sb|data)
Pr(none|data)
0.2000
0.2000
0.2000
0.2000
0.2000
70a
2.10961
add
Ch1
0.1402
0.4964
0.1317
0.2239
0.0077
70b
1.04151
add
Ch2
0.1543
0.6618
0.0781
0.1056
0.0002
70c
1.12237
add
Ch3
0.0700
0.8309
0.0436
0.0555
0.0000
70d
1.38835
add
Ch5
6%
90%
2%
2%
0%
68
Flat prior
SUMPRODUCT of prior and pdf
New probability normalised pdf x ptior / Sum

FIGURE 5

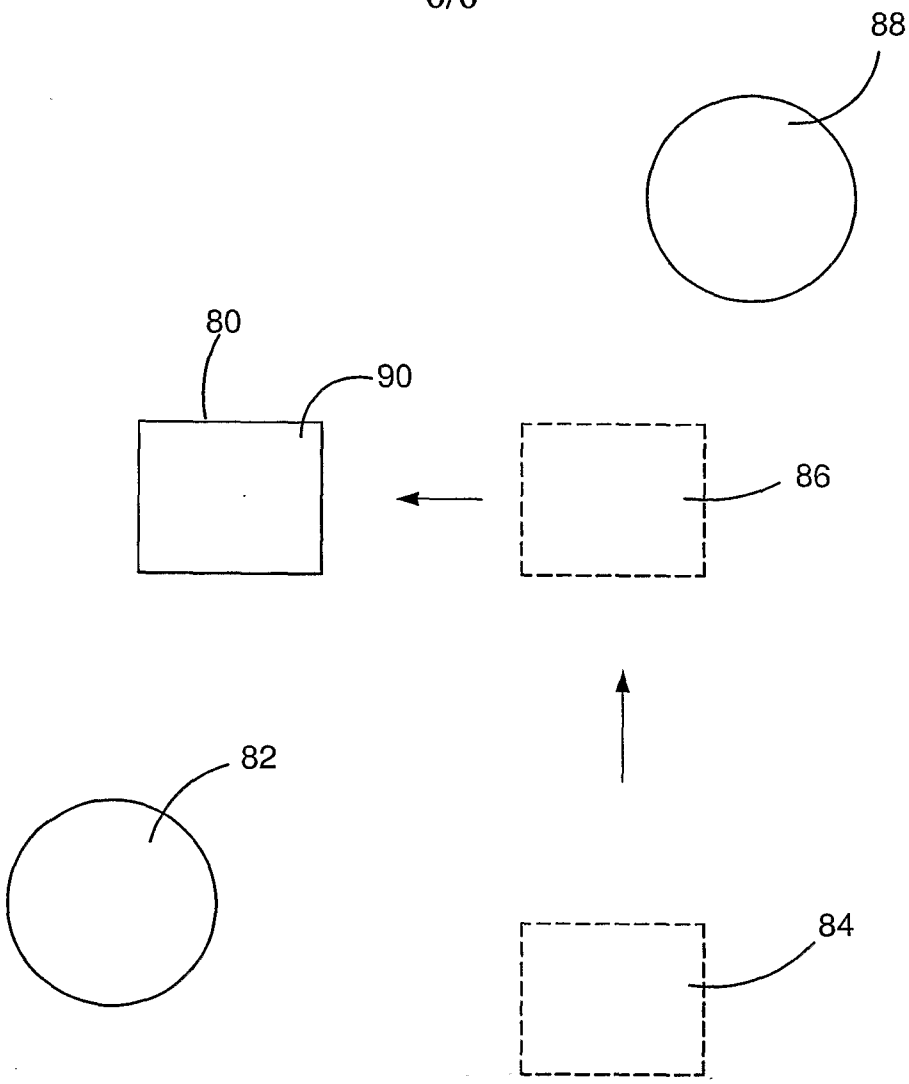


FIGURE 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2007/001214

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.

G06F 19/00 (2006.01)

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)

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 practicable, search terms used)
USPTO, EPO, WIPO, Google Scholar using IPC (where available) and keywords including electronic nose, sensor+, detector, bayes+, probab+, statistic+, source+, location and origin

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 7096125 B2 (PADMANABHAN et al) 22 August 2006 See whole document	1 to 35
A	US 6970560 B1 (HENCH et al) 29 November 2005 See whole document	1 to 35
A	US 5416724 A (SAVIC) 16 May 1995 See whole document	1 to 35
A	US 6496742 B1 (LABRECHE et al) 17 December 2002 See whole document	1 to 35

[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
14 September 2007	25 SEP 2007
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	Authorized officer **LUKE DAVESON** AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No : (02) 6283 2773

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 1065578 A1 (SIEMENS CORPORATE RESEARCH INC) 3 January 2001 See whole document	1 to 35
A	US 5832411 A (SCHATZMANN et al) 3 November 1998 See whole document	1 to 35

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2007/001214

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

SEE SUPPLEMENTARY SHEET

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Supplemental Box
(To be used when the space in any of Boxes I to VIII is not sufficient)

Continuation of Box No: III

This International Application does not comply with the requirements of unity of invention because it does not relate to one invention or to a group of inventions so linked as to form a single general inventive concept.

In assessing whether there is more than one invention claimed, I have given consideration to those features which can be considered to potentially distinguish the claimed combination of features from the prior art. Where different claims have different distinguishing features they define different inventions.

This International Searching Authority has found that there are different inventions as follows:

a. Claims 1 to 18 are directed to a method of determining the probability that sampled data are from a particular known source involving calculating probability density functions for each possible source and multiple sets of probabilities for each possible source. It is considered that the combination of calculations comprises a first special technical feature.

b. Claims 19 to 35 are directed to a method of determining the location of a sensor relative to a source which involves steps of obtaining data from the sensor at a plurality of locations and calculating the probability that the data obtained by a sensor is of a particular known source. It is considered that using the data obtained by a sensor from the multiple locations to determine the relative location of the sensor is a second special technical feature.

Although all claims define the common feature of calculating the probability that data has been derived from a particular source this feature is not novel in the light of at least US 5832411 A (SCHATZMANN et al) 3 November 1998 (see particularly column 9, lines 12 to 33). As the common feature is known it cannot be a special technical feature. Therefore the claims do not satisfy the requirement of unity of invention a posteriori.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/AU2007/001214

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
US	7096125	AU	2002368179	US	2004064260	US	2007093970
		WO	2004023413				
US	6970560	AU	14812/01	AU	15935/01	AU	15969/01
		AU	17598/01	AU	17623/01	AU	45059/01
		AU	52872/01	EP	1232569	EP	1232636
		US	6711598	US	6834109	US	6870901
		US	6970415	US	6978015	WO	0135607
		WO	0135608	WO	0135609	WO	0135611
		WO	0135614	WO	0139032	WO	0147170
US	5416724	US	5675506				
US	6496742	AU	90809/98	EP	1007961	FR	2767942
		WO	9912028				
EP	1065578	JP	2000356696	US	6785636		
US	5832411						

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

END OF ANNEX