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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 April 2003 (17.04.2003)

PCT

(10) International Publication Number
WO 03/032629 A1

(51) International Patent Classification[7]: H04N 5/232

(21) International Application Number: PCT/GB02/04458

(22) International Filing Date: 1 October 2002 (01.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0123921.9 5 October 2001 (05.10.2001) GB

(71) Applicant *(for all designated States except US)*: **HEWLETT-PACKARD COMPANY** [US/US]; 3000 Hanover Street, Palo Alto, CA 94304 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **GROSVENOR, David, Arthur** [GB/GB]; 18 Beaufort Road, Frampton Cotterell, Bristol BS36 2AD (GB).

(74) Agent: **LAWRENCE, Richard, Anthony**; Hewlett-Packard Limited, Intellectual Property Section, Filton Road, Stoke Gifford, Bristol BS34 8QZ (GB).

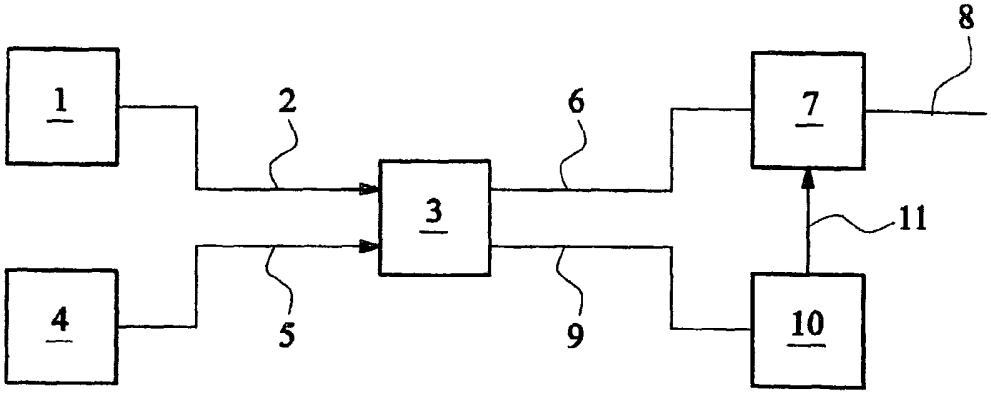
(81) Designated States *(national)*: JP, US.

(84) Designated States *(regional)*: European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR).

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*

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.

(54) Title: AUTOMATIC PHOTOGRAPHY



(57) Abstract: In automatic photographic or electronic camera apparatus, sounds from a predetermined person are preferentially received or identified to provide an audio signal which is then used to produce a saliency signal (11) (indicative of the person's response to circumstances) to control the camera 1 or an image signal (2, 6) produced therefrom. The sounds may be detected by a sound conduction microphone (4), e.g. on the head or throat of the person, or by voice identification circuitry used with a less specific microphone not necessarily mounted on the person. The person may or may not be wearing the camera. Control may be in real time, or the sound or saliency signal may be recorded (3) together with an image signal for subsequent control (7) of the latter. The audio signal may be categorised by waveform analysis, e.g. for the recognition of sounds such as speech (word recognition and/or prosody) or non-speech vocal sounds, and/or non-vocal sounds; signal amplitude may also be taken into account.

WO 03/032629 A1

Automatic Photography

The automation of photography has always been of interest, but has received increasing attention over the past few decades. Automatic exposure control and auto-focus are now standard features on most cameras, and many cameras have other electronically controlled or electrical features. Video and still cameras with electronic image sensors in the form of one or more photosensitive arrays are becoming increasingly available to the public, and are popular since they provide facilities not previously available, including the means to record images cheaply, to select edit and print only those images which are considered acceptable, and to do so at home using only a PC and ancillary equipment, and to transmit and receive images using e-mail.

However, in general, cameras bought off the shelf still require an operator to decide upon the composition of the actual picture to be taken, including decisions as to picture framing (direction and zoom) and timing. While the advent of electronic cameras does provide the ability to record a number of pictures in sequence, or even a video clip, from which a suitable image can then be selected, and while this can be followed by a degree of editing, the operator still has to point the camera in generally the correct direction and to press the operating button at around the correct time.

Since both still and video electronic cameras have sensors capable of providing a sequence of image signals (an "image sequence"), they will be broadly referred to herein as "electronic cameras", and their output will be referred to as an "image signal" irrespective of whether a still image or an image sequence results, and irrespective of whether the image signal is recorded at the camera or elsewhere. References to a picture or image should be read as covering both stills and image sequences. Unqualified references to a "camera" should be read as relating either to an electronic camera or a camera using a photosensitive recording medium, e.g. a film camera for still exposures or a cine camera (a "photographic" camera).

Camera controls such as autofocus and exposure relate more to the technical side of picture taking. These have little or nothing to do with picture composition, including framing and timing, but there is now increasing interest in this aspect and the

provision of intelligent still and video cameras which have a facility for picture taking or picture selection where conscious operator input is either further reduced or eliminated altogether.

One possible solution is to store the whole of the image signal from a continuously operated camera, or the whole of a sequence from a cine camera, i.e. an image sequence, and manually to select suitable single images or reduced image sequences therefrom at a later period. While this reduces the burden on a camera operator, or avoids the need for a camera operator altogether, not only does this approach require a very large amount of image signal or film storage, but if a human operator is later required to inspect the stored image sequence during the selection process it is also very time consuming and thus expensive. Therefore it is highly desirable either to have at least some degree of image selection before storage, or, if the whole image sequence is stored, to record therewith a further signal indicative of where the most interesting portions of the sequence are likely to be found.

Known types of automatic camera apparatus in which a degree of picture composition is effected encompass a broad spectrum. At one end lie forms of equipment in which the conditions for picture taking are strictly predetermined and extremely restricted, for example booths for taking passport photographs where the camera is operated with a fixed timing after a seated customer in an essentially fixed position has put money into the machine.

More freedom is permitted in apparatus where picture taking is automatically triggered in response to a predetermined condition or set of conditions, for example in response to the outcome of an algorithmic computation operating upon the output of one or more sensors. However, one problem with apparatus of this type is in defining the set of conditions necessary for picture taking to be enabled.

In certain circumstances, the prescribed conditions are physically predetermined. Security cameras which commence operation in response to an alarm input, and traffic speed cameras which respond to a vehicle being driven too fast, fall within this category. It is also known to take pictures of visitors on a ride at a theme park by

triggering a suitable located fixed camera which is actuated when the car in which the visitor is riding trips a switch, such as by cutting an infra-red beam. Such systems are disclosed in International Patent Application No. WO 98/10358 (Goldberg), European Patent Application No. 0 953 935 (Eastman Kodak), US Patent No. 5,655,053 (Renie) and US Patent No. 5,576,838 (Renie) for example.

Other systems allow rather more latitude, by recognition or detection of a feature associated with the feature to be imaged. Thus US Patent No. 5,844,599 (Hildin) discloses a video camera system for use with a group of people around a conference table, where each person wears a microphone, and the camera is arranged to respond by being directed towards the person who is speaking at the time, and to track the movements of that person. This is a system more intended for viewing than image recordal. European Patent Application No. 0 660 131 (Osen) describes a system where global positioning systems are employed to track a feature, such as an aircraft, so that cameras can be automatically directed thereto.

In this respect, use is increasingly being made of tags for identifying features being imaged/or for triggering the camera. This occurs in International Patent Application No. WO 98/10358 (Goldberg) mentioned above and in US Patent No. 5,694,514 (Evans) to enable pictures of different features to be easily identified. International Patent Application No. WO 98/10358 (Goldberg) describes a number of camera systems including the use of a tag to actuate a camera and to provide data for associating the image with the person wearing the tag; in a variant, facial recognition is used to identify the feature, working from an image and other relevant data initially fed into the system.

However, in comparison to a personal camera user who can employ a hand-held camera at any time and wherever they are to take pictures of anything considered to be of interest, in all of the prior art mentioned above there remain significant restrictions as to the manner of picture taking. Most known automatic camera apparatus either employs a fixed camera and/or has narrow limitations on the circumstances under which the camera is triggered, e.g. in response to a tag or a conditional or alarm input associated with a restricted condition or set of conditions.

In commercial installations such as theme parks, the cameras are essentially fixed and located to capture an image under specific circumstances, for example where it is anticipated that at least some of the stored images will prove to be sufficiently composed and appealing to be of interest to a would-be purchaser.

It would be desirable to be able to provide a camera which is portable and which can automatically capture different images irrespective of the person's location, as would be the case for a manually operated camera.

It could therefore be considered that a further desirable feature of an automatic camera apparatus is the provision of a signal (a "saliency" signal) indicative of when the viewed image may relate not be of a predetermined type but may nevertheless be interesting to a person, e.g. the camera user. This signal could be used to trigger selective camera operation or selective image storage in real time, or it could be used in conjunction with a continuously stored image signal in a later image selection process; if the signal has more than two values, i.e. indicative of the relative likelihood that the image relates to an interesting feature, then a combination of these techniques could be adopted, viz. storage of all signals possibly of interest whilst rejecting other signals, and later selection from the stored signals, with initial preference being given to those parts of the signal indicated as being more interesting. Such a feature is contemplated in respect of the present invention.

It is known that sounds coming from a person can be indicative of the person's reaction to the environment where the person is located. Such reactions may be to a particular event or to the general ambience of the environment. The reaction could be, for example, to a viewed or visible scene or event, to an interaction with another person, to sounds whether sudden (such as a bang) or continuous (such as music) and to other stimuli such as those experienced when partaking in an event or when enjoying entertainment such as a fairground ride. Whatever the reaction, when it occurs the scene viewed by the person may either represent the cause of the reaction, or at least be evocative, at a later date, of the cause of the reaction. Accordingly, in embodiments of the present invention the sounds emitted from a person are employed in the production of a saliency signal. The invention finds particular use with cameras

which are portable or carried by a person, but is not restricted thereto, and may be used with cameras which are mounted on inanimate objects, which can move, or are mounted at fixed locations, or which follow a fixed track

In a first aspect the present invention provides imaging apparatus comprising an electronic camera providing an image signal, sound detection means capable of preferentially receiving or identifying sounds from a predetermined person and producing an audio signal in response thereto, and recording means for recording the audio signal together with the image signal. In the first aspect, the invention may further comprise means for playing back the audio and image signals, audio means for providing a saliency signal in response to the replayed audio signal, and control means responsive to the saliency signal for controlling the replayed image signal.

In a second aspect the invention provides imaging apparatus comprising a camera, camera, sound detection means capable of preferentially receiving or identifying sounds from a predetermined person and producing an audio signal in response thereto, and audio means for providing a saliency signal in response to the audio signal, and control means responsive to the saliency signal for controlling the camera.

In the second aspect, preferably the camera is an electronic camera providing an image signal, but it could be a conventional still or cine film camera. Control of the camera may be actuation for taking a picture.

In a third aspect the invention provides an electronic camera providing an image signal, sound detection means capable of preferentially receiving or identifying sounds from a predetermined person and producing an audio signal in response thereto, audio means for providing a saliency signal in response to the audio signal, and control means responsive to the saliency signal for controlling the image signal.

The invention extends to methods of imaging using apparatus according to the first to third aspects.

As particularly described later, a preferred sound detection means is a sound conduction microphone adapted to be fitted to the head of a wearer (the

"predetermined person") for preferentially detecting sounds emitted from the wearer's head and/or throat and producing an audio signal in response thereto. However, other techniques are known for preferentially receiving or identifying sounds from a predetermined person, and use thereof in lieu of the use of a sound conduction microphone falls within the scope of the present invention. These techniques include the use of voice recognition software acting on a microphone capable of receiving sounds from more than one person, for example – such software may be trained to recognise speech, in particular, from a specific user. Nevertheless, the use of a sound conduction microphone can often offer a particularly simple and effective way of performing the intended function.

Sound conduction mounted microphones, which are common to all three aspects of the invention are well known, and normally include a vibration transducer for detecting sound vibrations conducted through the wearer's body, the bone(s) of the skull for head and ear mounted microphones. They are differentiated from microphones which pick up sound transmitted through the air.

Typical examples of ear mounted microphones are described in International Patent Application No. WO 86/07157 (ACR Electronics); US Patent No. 4,588,867 (Konomi) and US Patent No. 5,692,059 (Kruger). Other skull mounted microphones are secured to the top of the head. All of these include a vibration transducer for picking up sounds directly transmitted through the head, normally via bone conduction. Another typical form of head mounted microphone is mounted in physical contact with the throat. Sounds transmitted through the air may also be sensed by these microphones. Circuits for operating such microphones are disclosed for example in International Patent Application No. WO 00/21194 ((Resound Corporation) and in European Patent Application Serial No. 0 984 660 (Nippon Telephone and Telegraph Corporation).

Cameras which are activated by a sound signal are known. Examples are disclosed in US Patents No 5,014,079 and 5,107,288 both to Konica Corporation; and US Patent No. 5,162832 (Konishoroku Photo Ind.), but in general they are arranged to respond only to one particular sound feature, for example the envelope of intensity (or sound

pressure) in these particular cases, and the microphone is located at the camera and is not a sound conduction microphone.

Also known from US Patent No. 5,844,599 (Lucent Technologies Inc.) is an automatic camera system for use at conferences where each participant is provided with a microphone which acts as a tag when that person is speaking, the system being arranged so that the camera is automatically pointed at the speaker, or the loudest speaker. Although this system involves personal microphones, they are not specifically sound conduction microphones, and the only sound parameter which is utilised is the sound intensity.

None of this prior art is directed to the use of an audio signal from a head or throat mounted sound conduction microphone for control of the operation of a camera or the control of an image signal produced therefrom.

Systems also exist for the classification of sounds, and the audio means mentioned above may comprise audio analysis means for analysing the waveform of the audio signal to categorise it as falling within one of a set of predetermined categories.

US Patent No. 5,712,953 (Langs) discloses a system for classifying an audio signal as music or non-music, and refers to other such prior art. Perhaps more pertinently for the present invention, US Patent No. 5,918,223 (Muscle Fish) discloses a system for the more detailed classification of audio signal by comparison with given sound signals. The system is said to be useful in multimedia database applications and Internet search engines for finding audio data files which sound similar to the given sounds. One given example is the identification of applause and cheers in a soundtrack. Classes of sounds such as "bird sounds" or "rock music" can be created by the user.

In a basic form of the apparatus of the invention the audio analysis means may be adapted to recognise only one particular form of sound, such as laughter, emitted by the predetermined person, and in such a case the most basic set of categories will consist of "laughter" and "other".

However, it is possible to categorise a number of sounds emitted by the predetermined person, albeit some with more difficulty than others, thus to enlarge the range of identifiable categories and/or events to which the categories relate. Sounds from the head of a predetermined person can be split into non-speech and speech, which need to be considered separately.

Typical non-speech sounds include vocal sounds such as laughter, crying, humming, coughing and sneezing, and non-vocal sounds such as sniffing, scratching, eating, teeth grinding (bruxism) and eructation.

Speech itself can be split into words and the manner in which they are spoken (prosody). Word recognition techniques are sufficiently advanced for many words to be accurately recognised. Prosody includes such features as the voice intensity pattern (raised or lowered voice, or a pattern of variation such as is associated with the reading of a string of sentences or a poem, the asking of a question, or the making of a statement; hesitation; the manner in which words are spoken, i.e. the different stresses associated with syllables; the frequencies in the voice; speaking rate; the duration of parts of speech. It consists of a phonological aspect, identifying discrete abstract units of speech, and a phonetic aspect relating to continuously varying acoustic information associated with such units, and may reflect the mood of the predetermined person, e.g. excitement, anger or relaxation.

Some sounds, such as singing may be considered as falling into both speech and non-speech associations.

In the invention the categorisation performed by the audio analysis means may additionally be a function of the overall signal (envelope) amplitude, although this may be considered to be part of prosody in any case where speech is involved.

Thus the audio analysis means may be arranged for the recognition of vocal sounds, and for recognising whether the sounds are associated with speech or non-speech. Where speech is involved, the audio analysis means may be arranged for word recognition and/or for the recognition of features associated with prosody. However, the audio analysis means may alternatively or additionally be arranged for the

recognition of non-speech vocal sounds and or for the recognition of non-vocal sounds.

In addition, the audio analysis means may be arranged to take account of the overall signal amplitude or envelope when assigning a category to the audio signal. For example it may be arranged to differentiate between loud laughter and quiet laughter.

The audio analysis may be provided with a predetermined series of reference waveforms against which to perform the categorisation. Alternatively or additionally, and preferably, it may be arranged so that it can be trained to recognise new sounds so that it can be adapted to the specific mannerisms of the predetermined person.

Word recognition may be useful for voluntary and involuntary camera or image signal control. Words and expressions such as "Hello", "Goodbye", Wow" and "Go on" may be useful for involuntary initiation of picture taking when dealing with people or attending an event, whereas a command such as "Take that" may be recognised as a deliberate command to take a picture. Words such as "Yes" when spoken with a sense of triumph, or "No" when spoken as in disbelief are examples of word recognition combined with prosody. Laughter and humming may be taken as two examples of non-speech vocal sounds which could initiate picture taking or selection. Loud laughter may unconditionally effect such initiation, whereas softer laughter may need to be combined with a further conditional signal to determine whether or not this is effected.

A saliency signal may be developed according to the categorisation thus performed. In it simplest, and possible most practical form, the saliency signal is binary, so that a "yes" or "no" signal is issued at any time in response to the sound from the microphone. This may be utilised alone or in logical combination with other control signals for use as a control signal. The latter may be used for example for controlling the camera or the image signal, for example by telling the camera to take a picture or to commence an imaging sequence, or by operating a gate for selective transmission display or recording of the image signal.

Alternatively the saliency signal may have a range of discrete or continuous values. This may be useful for someone scanning through a complete image sequence, either for identifying particular sounds associated with a particular value of saliency signal, or for searching for the most interesting sounds as defined by a predetermined ordering of values assigned to different sounds. Also, in such a form it could be numerically combined, for example according to a predetermined algorithm, with other signals for producing a control signal of the type mentioned in the preceding paragraph, possibly after thresholding the resulting signal.

To derive a numerical saliency signal, different recognised sounds (and this could include similar sounds having different intensities as being different recognised sounds) may be assigned different values. This may be sufficient, or where intensity has not already been taken into account, the assigned value may be modified continuously according to the intensity of the sound.

In a preferred form of the invention the camera is worn by the predetermined person, preferably a sound conduction microphone wearer, and is preferably worn so as to point forwards of the wearer. While the camera could be chest mounted, e.g. as a badge, it is particularly preferred that the camera be provided with fastening means for locating it, or at least the imaging optics thereof, to the head of the person, for example in the frame for a pair of glasses or in headgear such as a hat. In this way the camera direction more accurately follows the gaze direction of the wearer. If appropriate, the field of view of the camera may be made significantly greater than that of the human wearer, possibly to cover the full range of eye movement. This is particularly advantageous if the number of available pixels, e.g. of a sensor array in a solid state camera, is correspondingly increased so as to maintain the resolution for a view approximating the normal human field of view.

For true independent portability in the first aspect of the invention the predetermined person is also provided with a recorder for recording both the image and audio signals, e.g. for later development of the saliency signal for controlling a replayed image signal, or for indicating parts of a replayed image signal which could be of interest to the viewer.

For true independent portability in the second and third aspects of the invention the predetermined person may also be provided with the audio means for providing the saliency signal. This may then be recorded alongside the complete image signal, e.g. for use as mentioned in the preceding paragraph, or for control of the image signal, such as by gating it to a display or recorder, or for camera control or actuation.

However, any of the audio, saliency and image signals may be transmitted to a remote location if appropriate. One occasion when transmission of the audio or saliency signal is appropriate is when the camera is not worn by the predetermined person. For example, the apparatus may be part of an installation where the camera is fixed for viewing a predetermined scene, and the person then acts as a photographer voluntarily (word commands) or involuntarily (non-speech sounds) controls operation of the camera, or gating of the image signal, or at least indicates the saliency of a recorded signal. This could be of use at a special occasion such as a birthday party. It should be noted that the camera need not be fixed, and could for instance follow a predetermined track as in a fairground ride, where the predetermined person is tracked. Other suitable arrangements, and uses of the invention, will no doubt come to the mind of the reader.

While a necessary part of preferred embodiments of the invention is the use of a sound conduction microphone, it should be noted that one or more further microphones may be provided, and that at least one such further microphone may be a "conventional" microphone for receiving sound via the air.

The signal from the "conventional" microphone may be recorded, for example with the image and/or sound conduction microphone signal for replaying at a later date. In this respect it should be borne in mind that it would be more normal to replay the signal from a "conventional" microphone for listening purposes, the signal from the conduction microphone giving rise to a rather unreal sound for this purpose.

Alternatively or additionally, it may be combined with the signal from the sound conduction microphone. For example it may be possible to subtract a significant part of the sound produced by a person wearing a sound conduction microphone from the

signal produced by the "conventional" microphone, leaving what may be considered to be sounds ambient to the sound conduction microphone wearer. This would also allow an inference to be made as to whether the wearer of the conduction microphone was involved in a conversation with another party, as by analysing or comparing the pattern of the ambient sound and that detected by the conduction microphone. This could be a feature determine the saliency signal, i.e. the latter may be derived by analysis of the outputs of a combination of microphones, and such combination may include a "conventional" microphone.

The camera may be capable of capturing still pictures and/or image sequences. It may include a "conventional" microphone as referred to in the preceding paragraph.

The reader will understand that the audio analysis may be performed, in most cases more efficiently and effectively, by the use of Fourier techniques which are well known per se.

Further features and advantages of the invention will become clear upon a consideration of the appended claims, to which the reader is referred, and upon reading the following more detailed description of a preferred embodiment of the invention made with reference to the accompanying drawings in which Figures 1 to show first to fourth embodiments of the invention in schematic form and Figure 5 exemplifies an audio analysis in outline schematic form.

Where appropriate, like reference numbers denote similar items in each of the drawings.

In Figure 1, a head mounted electronic camera 1 produces an image signal 2 for storage in a multi-channel recorder 3. A sound conduction ear mounted microphone 4 produces an audio signal 5 for synchronous recording with the image signal at recorder 3. Subsequently, during playback, the replayed audio signal 9 is passed to a sound categorisation circuit 10, which produces a binary (saliency) signal 11 according to whether or not the sound waveform shape approximates any one of a number of stored waveform shapes. Signal 11 is used to operate a gate 7 for selective transmission 8 of the replayed image signal 6. In a variant, the signal 11 is used to

operate a signal to a human viewer of the replayed and displayed image signal 6 to indicate that particular attention should be paid thereto.

In Figure 2, the signal 5 from a sound conduction skull mounted microphone 4 is categorised in circuit 10 and a change in the binary (saliency) output 11 in one direction is used to operate a photographic or electronic camera 1, to cause it to take a single picture or to commence an image sequence (a video or cine clip). In the latter case, the clip continues until signal 11 changes state in the other direction. Also, in the case of an electronic camera 1 the image signal 2 is recorded or displayed at 12.

In Figure 3, the output 5 of a throat mounted sound conduction microphone 4 is categorised in circuit 10 to produce a saliency signal 13 for control of a gate 7 selectively transmitting the image signal 2 to a recorder or display 12. The signal 13 may be a binary signal and may be directly coupled to the gate 7, but as shown, it is combined with one or more other signals 16 in a circuit 14 which produces a binary signal 15 for controlling the gate 7.

In a first form of this circuit the signal 13 and the signal(s) 16 are binary signals and they are logically combined in the circuit 14. For example the signal 13 may be ANDed with a signal 16 indicating that the picture composition (obtained by spatial segmentation of the image signal 2) is acceptable, and the result may be ORed with a signal 16 instructing the camera to take a picture in any case. The latter signal could in fact be a signal from the circuit 10 on detection of a command such as "take that", but is not necessarily so.

In a second form of this circuit, the circuit 10 assigns values (the saliency signal) to the audio signal 5 according to the type of sound which has been detected, and these values are combined in an algorithm circuit 14 with numerical signals 16 from other sources according to a predetermined algorithm, the output being thresholded to provide the binary signal 15. In this case, an over-riding command such as "take that" will be assigned an extreme value so as to determine the value of the signal 15 irrespective of the values of the signals 16.

In Figure 4 the signal 5 from ear mounted sound conduction microphone 4 is transmitted via a wireless coupling 18, 19, e.g. via rf, infra-red or induction, to a multi-channel recorder 3 which also receives the image signal output 2 of a fixed camera 17, and control signal(s) 16 from at least one other source. Subsequently the signals are replayed as image signal 6 and control signals 20, the latter being processed by a circuit 14 and used to control image signal gate 7 as in Figure 3.

In a variant of Figure 4, the camera is fixed to a vehicle in a fairground ride and observes the microphone wearer during the ride. One of the signals 16 is an enable signal provided for a time period including that when the ride is in progress, and a short period before and afterwards.

Figure 5 shows in schematic form an audio analysis arrangement for providing a saliency signal 27. The audio signal 5 is received by a waveform shape analysis circuit 21 and an envelope amplitude analysis circuit 25.

The circuit 21 determines the type of waveform and produces one of a set of four signals 29 to 32 according to whether the type is determined to be vocal speech, vocal non-speech, non-vocal or indeterminate respectively. Each of the signals 29 to 31 is used to enable a respective circuit 22 to 24 to determine the degree of conformity of the waveform shape with each of a respective set of stored waveforms. The operative one of circuits 22 to 24 then proceeds to produce a label signal 33 to 35 uniquely identifying the stored waveform with which the incoming waveform most closely corresponds, providing there is sufficient correspondence, and otherwise provides a label signal signifying no correspondence, which equates to the signal 32 from circuit 21.

The signals 33 to 35 are combined and used with a digitised volume/amplitude signal 36 from the circuit 25 to address a two-dimensional look-up table. For a binary output 27 from the table, all of the table contents are either zero or one, but for a numerical saliency output, the table may contain a range of numbers.

Where in the absence of any indication of the overall waveform the determination of the envelope amplitude is a problem, circuits 21 and 25 are coupled to facilitate the

determination of envelope amplitude after the waveform shape has been determined, as shown by the dotted line 28. Alternatively or additionally, the outputs 33 to 35 can be coupled to the circuit 25.

In addition, where prosody is important, the output 33 of circuit 22 together with the signal 5 may be coupled to a prosody determining circuit for providing an alternative numerical output for inputting to the table 26, the numerical output in this case being indicative of the word and the associated prosody.

CLAIMS

1. Automatic imaging apparatus comprising an electronic camera providing an image signal, sound detection means capable of preferentially receiving or identifying sounds from a predetermined person and producing an audio signal in response thereto, and recording means for recording the audio signal together with the image signal.

2. Apparatus according to claim 1 wherein said sound detection means comprises a sound conduction microphone adapted to be worn on the head or throat region of a said predetermined person for preferentially detecting sounds emitted from the wearer's head and producing said audio signal in response thereto.

3. Apparatus according to claim 1 and further comprising means for playing back the audio and image signals, audio means for providing a saliency signal in response to the replayed audio signal, and control means responsive to the saliency signal for controlling the replayed image signal.

4. Automatic imaging apparatus comprising a camera, sound detection means capable of preferentially receiving or identifying sounds from a predetermined person and producing an audio signal in response thereto, and audio means for providing a saliency signal in response to the audio signal, and control means responsive to the saliency signal for controlling the camera.

5. Apparatus according to claim 4 wherein the camera is an electronic camera providing an image signal.

6. Apparatus according to claim 4 wherein the camera is arranged for image recordal on a photosensitive medium.

7. Automatic imaging apparatus comprising an electronic camera providing an image signal, sound detection means capable of preferentially receiving or identifying sounds from a predetermined person and producing an audio signal in response thereto, audio means for providing a saliency signal in response to the

audio signal, and control means responsive to the saliency signal for controlling the image signal.

8. Apparatus according to any one of claims 3 to 7 wherein said sound detection means comprises a sound conduction microphone adapted to be worn on the head or throat region of a said predetermined person for preferentially detecting sounds emitted from the wearer's head and producing said audio signal in response thereto.

9. Apparatus according to any one of claims 3 to 8, wherein the audio means comprises audio analysis means for analysing the waveform of the audio signal to categorise it as falling within one of a set of predetermined categories.

10. Apparatus according to claim 9 wherein said audio analysis means is arranged for the recognition of vocal sounds.

11. Apparatus according to claim 10 wherein said audio analysis means is arranged for the recognition of sounds associated with speech.

12. Apparatus according to claim 11 wherein said audio analysis means is arranged for word recognition.

13. Apparatus according to claim 11 or claim 12 wherein said audio analysis means is arranged for the recognition of features associated with prosody.

14. Apparatus according to any one of claims 9 to 13 wherein said audio analysis means is arranged for the recognition of non-speech vocal sounds.

15. Apparatus according to any one of claims 9 to 14 wherein said audio analysis means is arranged for the recognition of non-vocal sounds.

16. Apparatus according to any one of claims 9 to 15 wherein the audio analysis contains a series of reference waveforms against which to perform the categorisation.

17. Apparatus according to any one of claims 9 to 16 wherein the audio analysis is arranged so that it can be trained to recognise new sounds.

18. Apparatus according to any one of claims 9 to 17 wherein the audio analysis means is arranged to take account of the overall signal amplitude or envelope when assigning a category to the audio signal.

19. Apparatus according to any one of claims 9 to 18 a further comprising a convention microphone for receiving sounds via the air.

20. Apparatus according to claim 19 and including means for recording the signal from the conventional microphone.

21. Apparatus according to claim 19 or claim 20 wherein the audio analysis means is arranged to take account of the signal from the conventional microphone when categorising categorise said audio signal as falling within one of a set of predetermined categories.

22. Apparatus according to any one of claims 9 to 21 wherein the audio analysis means is arranged to provide a saliency signal having a value determined at least in part by the assigned category of the audio signal.

23. Apparatus according to claim 22 wherein the audio analysis means provides a binary said saliency signal according to the assigned category of the audio signal.

24. Apparatus according to claim 23 wherein the audio analysis means provides a saliency signal having one of a plurality of values according to the assigned category of the audio signal.

25. Apparatus according to claim 22 wherein the audio analysis means is arranged to provide a saliency signal which is discretely variable as a function of the assigned category and continuously variable as a function of the overall signal amplitude or envelope of the audio signal.

26. Apparatus according to claim 3 or claim 7, or any preceding claim as directly or indirectly dependent on claim 3 or claim 7, wherein said control means comprises a gate for selectively transmitting the image signal.

27. Apparatus according to claim 26 wherein said control means includes combination means for combining said saliency signal with other control signals to produce a gate control output for controlling said gate.

28. Apparatus according to claim 26 and claim 23 wherein said gate is arranged to be controlled directly by said binary saliency signal.

29. Automatic imaging apparatus according to any preceding claim wherein said camera is provided with fastening means for fastening it to a said predetermined person.

30. Automatic imaging apparatus according to any one of claims 1 to 28 wherein said camera is provided with fastening means for fastening at least the imaging optics thereof to the head of a said predetermined person.

31. Apparatus according to claim 7, or any one of claims 8 to 30 as directly or indirectly dependent on claim 7, and including recording means for recording the controlled image signal.

32. Apparatus according to claim 1, or any claim dependent thereon, or claim 31, wherein the recording means is provided with securing means for securing it to a said predetermined person.

33. Apparatus according to claim 1, or any claim dependent thereon, wherein said recording means is located at a recorder location remote from sound detection means, and comprising means for wirelessly transmitting said audio signal from said sound detection means to said recorder location.

34. Apparatus according to any one of claims 3 to 7, or any one of claims 8 to 30 as dependent therefrom, wherein said audio means is mounted at an audio means

location remote from the sound detection means, and comprising means for wirelessly transmitting said audio signal to said audio means location.

35. Apparatus according to any one of claims 3 to 7, or any one of claims 8 to 30 as dependent therefrom, wherein said control means is located at a control location remote from said audio means, and comprising means for wirelessly transmitting said saliency signal from said audio means to said control location.

36. Apparatus according to any one of claims 33 to 35 wherein said camera is mounted at a camera location other than on the said predetermined person.

37. A method of imaging comprising the steps of operating an electronic camera to produce a image signal, preferentially receiving or identifying sounds from a predetermined person and producing an audio signal in response thereto, and recording the audio signal together with the image signal.

38. A method according to claim 37 wherein said step of selectively identifying is performed using a sound conduction microphone fitted to the head or throat region of said predetermined person for preferentially detecting sounds emitted from the person's head.

39. A method according to claim 37 and further comprising the step of playing back the audio and image signals, producing a saliency signal in response to the replayed audio signal, and controlling the replayed image signal in response to the saliency signal.

40. A method of imaging comprising the steps of operating a camera to produce a image signal, preferentially receiving or identifying sounds from a predetermined person and producing an audio signal in response thereto, producing a saliency signal in response to the audio signal, and controlling the camera in response to the saliency signal.

41. A method according to claim 40 wherein the camera is an electronic camera providing an image signal.

42. A method of imaging comprising the steps of operating a camera to produce an image signal, preferentially receiving or identifying sounds from a predetermined person and producing an audio signal in response thereto, producing a saliency signal in response to the audio signal, and controlling the image signal in response to the saliency signal.

43. A method according to any one of claims 39 to 42 wherein said step of preferentially receiving or identifying is performed using a sound conduction microphone fitted to the head or throat region of said predetermined person for preferentially detecting sounds emitted from the person's head.

44. A method according to any one of claims 39 to 43 wherein the step of producing said saliency signal comprises analysing the waveform of the audio signal to categorise it as falling within one of a set of predetermined categories.

45. Automatic imaging apparatus substantially as hereinbefore described with reference to the accompanying drawings.

46. A method of imaging substantially as hereinbefore described with reference to the accompanying drawings.

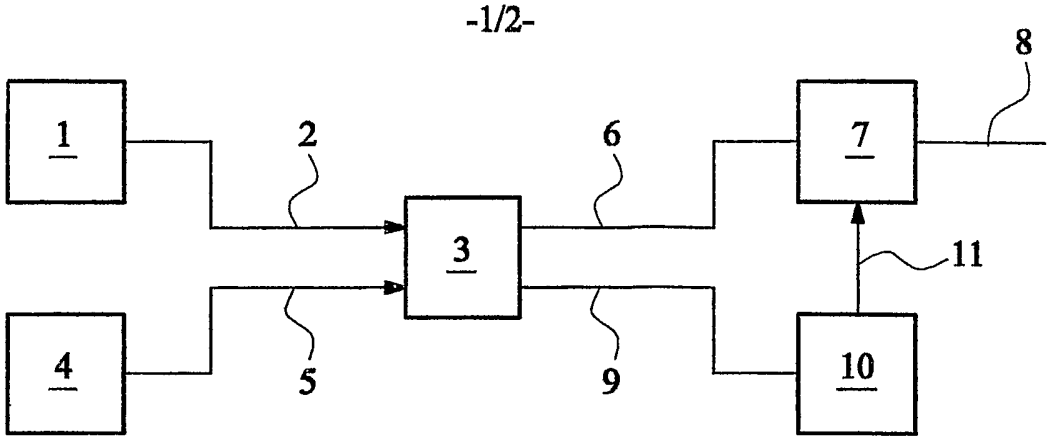


FIG. 1

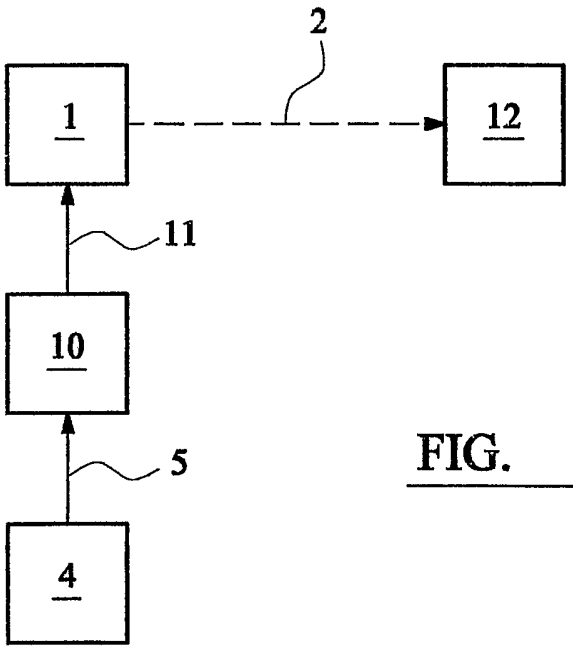


FIG. 2

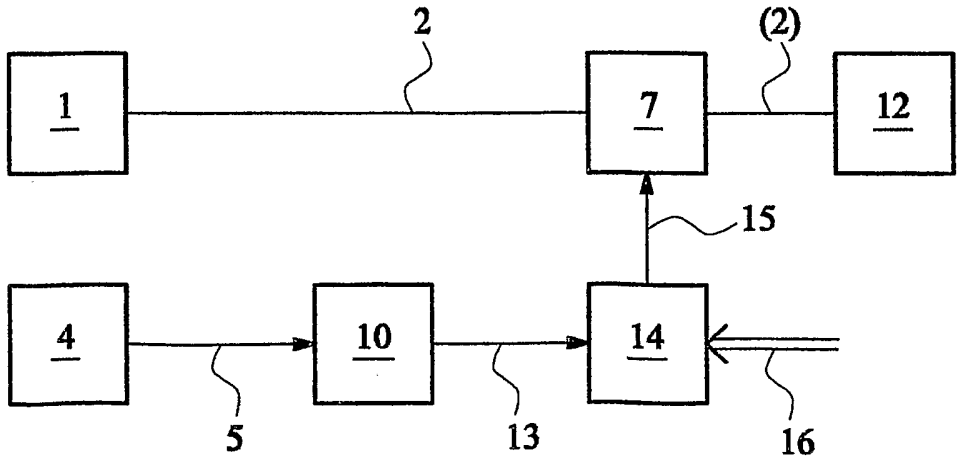


FIG. 3

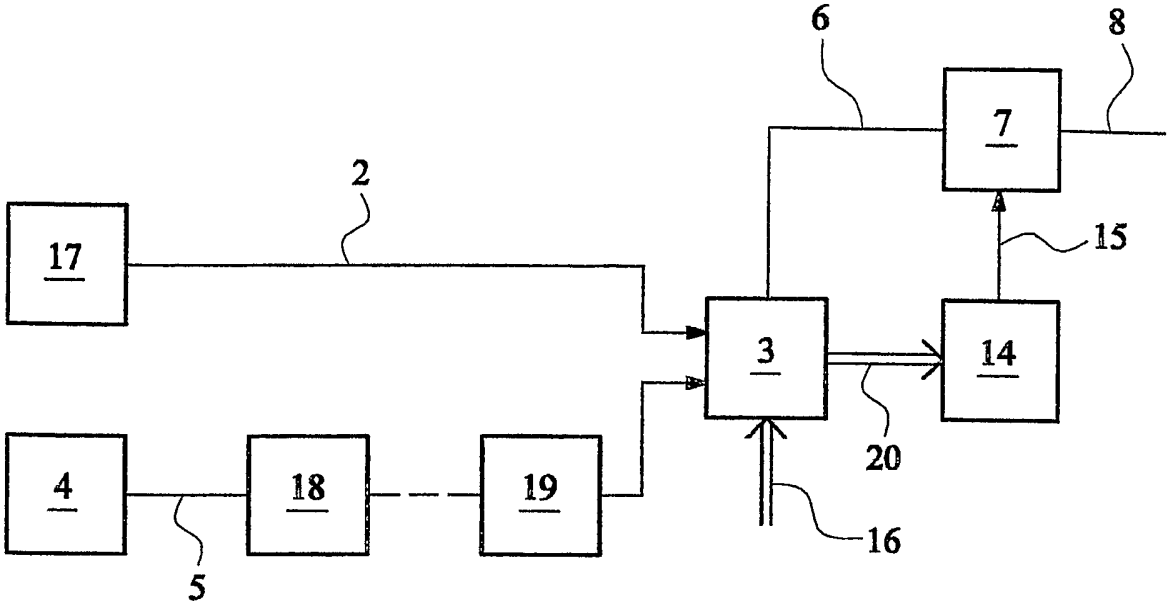

6
8
7
2
17
15
3
14
20
4
18
19
16
5

FIG. 4

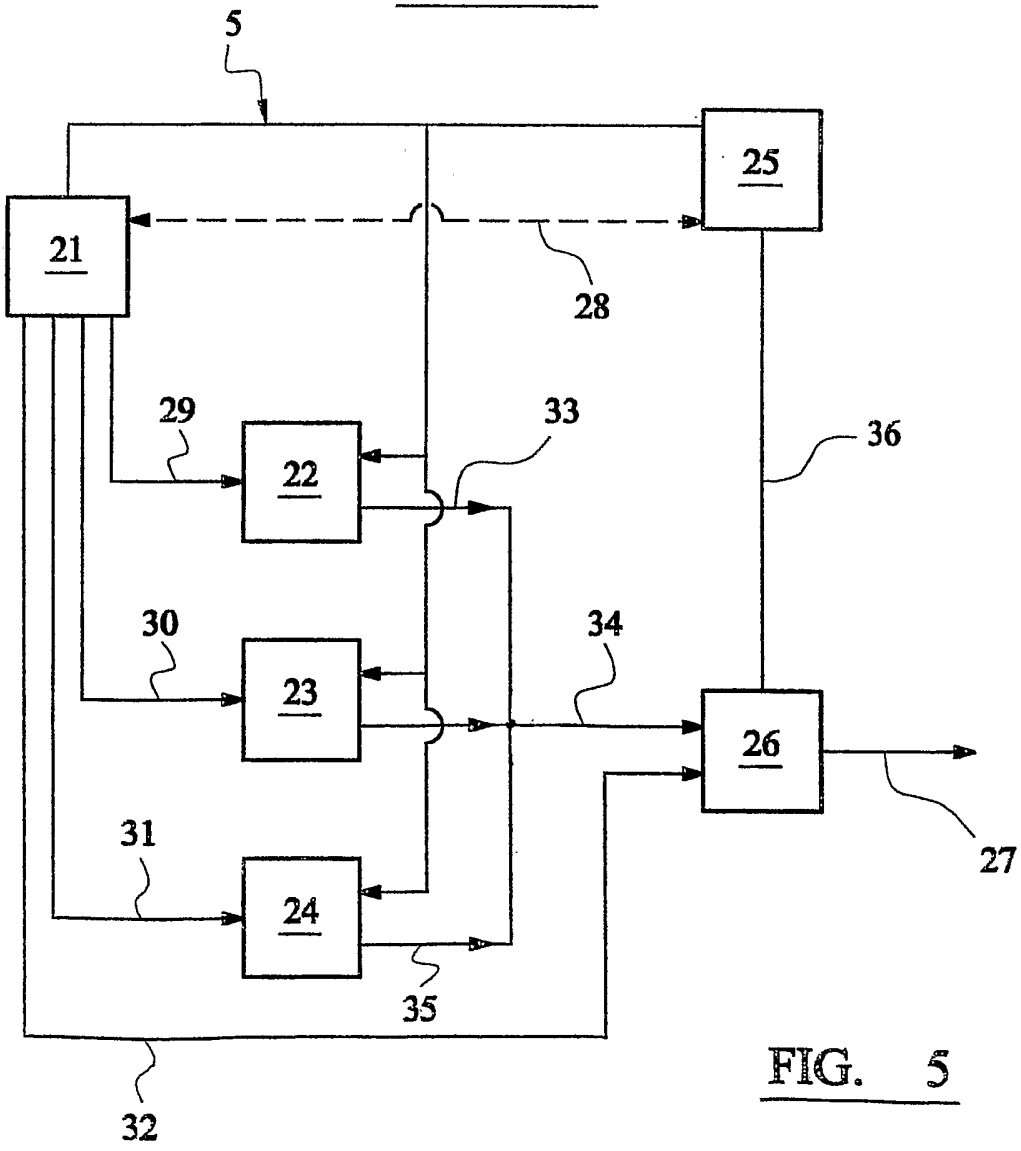

5
25
21
28
29
33
36
22
30
34
23
26
27
31
24
35
32

FIG. 5

INTERNATIONAL SEARCH REPORT

PCT/GB 02/04458

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N5/232

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 H04N G10L

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, 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.
X	WO 97 08896 A (SCIENTIFIC ATLANTA ;RAMSAY LEIGH WILLIAM (US); DALRYMPLE DALE BRIA) 6 March 1997 (1997-03-06)	1,4-7,9, 14-16, 18-24, 26,28, 31,33, 34,36, 37, 40-42, 44-46
A	page 2, line 19 -page 3, line 1 page 4, line 27 -page 5, line 33 page 7, line 26 -page 11, line 7 page 16, line 13 -page 17, line 31 page 19, line 9 -page 23, line 8 figures 1,12,14B	25,27,35

-/--

[X] Further documents are listed in the continuation of box C.

[X] 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	Date of mailing of the international search report
13 February 2003	06. 03. 2003

Name and mailing address of the ISA	Authorized officer
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	Didierlaurent, P

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 289 140 B1 (OLIVER THOMAS C) 11 September 2001 (2001-09-11) column 1, line 59 -column 6, line 14 column 8, line 60 - line 67 figure 1	1,4,5,7, 34,37, 40-42, 45,46
X	US 5 548 335 A (MITSUHASHI YASUO ET AL) 20 August 1996 (1996-08-20) column 6, line 38 -column 7, line 5 column 7, line 59 -column 8, line 48 figure 9	4,6, 9-12,15, 18,19, 21,22, 26,28, 29,31, 32,34, 40,42, 44-46
X	WO 01 54399 A (KONINKL PHILIPS ELECTRONICS NV) 26 July 2001 (2001-07-26)	4,5,7, 40-42, 45,46
A	page 3, line 1 -page 6, line 24 page 8, line 26 -page 12, line 13 figure 1A	1,8-28, 36,37,44
A	US 5 774 851 A (OHORA YASUNORI ET AL) 30 June 1998 (1998-06-30) column 11, line 16 -column 15, line 34	9,44
A	EP 0 984 660 A (NIPPON TELEGRAPH & TELEPHONE) 8 March 2000 (2000-03-08) cited in the application page 2, line 1 -page 3, line 10	2,8,38, 43
A	US 4 951 147 A (AKNAR ATILA ET AL) 21 August 1990 (1990-08-21) column 4, line 22 -column 5, line 23 figure 3	1

INTERNATIONAL SEARCH REPORT

Information on patent family members

PCT/GB 02/04458

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9708896 A	06-03-1997	WO 9708896 A1	06-03-1997
US 6289140 B1	11-09-2001	US 6226422 B1	01-05-2001
		CN 1258162 A	28-06-2000
		EP 1014338 A1	28-06-2000
		JP 2000215021 A	04-08-2000
		KR 2000057082 A	15-09-2000
		TW 442772 B	23-06-2001
		US 6295391 B1	25-09-2001
US 5548335 A	20-08-1996	US 5579046 A	26-11-1996
		DE 4124927 A1	06-02-1992
		GB 2247805 A ,B	11-03-1992
		JP 2687712 B2	08-12-1997
		JP 4175073 A	23-06-1992
		US 5521635 A	28-05-1996
WO 0154399 A	26-07-2001	WO 0154399 A2	26-07-2001
		EP 1186162 A2	13-03-2002
US 5774851 A	30-06-1998	JP 62039900 A	20-02-1987
		JP 1952618 C	28-07-1995
		JP 6077198 B	28-09-1994
		JP 62145299 A	29-06-1987
		JP 62145298 A	29-06-1987
EP 0984660 A	08-03-2000	JP 7312634 A	28-11-1995
		JP 3082825 B2	28-08-2000
		JP 8070344 A	12-03-1996
		EP 0984660 A2	08-03-2000
		EP 0984661 A2	08-03-2000
		CA 2149563 A1	19-11-1995
		DE 69525987 D1	02-05-2002
		DE 69525987 T2	19-09-2002
		DE 69527731 D1	12-09-2002
		EP 0683621 A2	22-11-1995
		US 5933506 A	03-08-1999
US 4951147 A	21-08-1990	AT 105431 T	15-05-1994
		AU 597134 B2	24-05-1990
		AU 8279587 A	16-06-1988
		WO 8804082 A1	02-06-1988
		BR 8707544 A	14-03-1989
		DE 3789771 D1	09-06-1994
		DK 416988 A	25-07-1988
		EP 0290548 A1	17-11-1988
		FI 883448 A	21-07-1988
		GR 871786 A1	09-03-1988
		HU 59761 A2	29-06-1992
		IE 873099 L	25-05-1988
		JP 1501514 T	25-05-1989
		KR 9403702 B1	27-04-1994
		MC 1945 A	19-05-1989
		NZ 222628 A	27-11-1990
		ZA 8708650 A	18-05-1988

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 02/04458

Box I Observations where certain claims were found unsearchable (Continuation of item 1 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 II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

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

see additional sheet

As a result of the prior review under R. 40.2(e) PCT, no additional fees are to be refunded.

1. [X] 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 an additional fee, this Authority did not invite payment of any additional fee.

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

[X] The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/GB 02/04458

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-3,8-39,43-46

   Automatic imaging apparatus and method comprising an electronic camera, sound detection means and recording means

2. Claims: 4-6,8-36,40,41,45,46

   Automatic imaging apparatus and method comprising a camera, sound detection means and audio means capable of providing a saliency signal and control means responsive to the saliency signal for controlling the camera

3. Claims: 7-36,42-46

   Automatic imaging apparatus and method comprising a electronic camera, sound detection means and audio means capable of providing a saliency signal and control means responsive to the saliency signal for controlling the image signal