

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2005-506003
(P2005-506003A)

(43) 公表日 平成17年2月24日(2005.2.24)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
HO4N 5/225	HO4N 5/225	5C122
G10L 15/00	HO4N 5/232	5D015
G10L 15/10	G10L 3/00	551G
HO4N 5/232	G10L 3/00	531N
// HO4N 101:00	HO4N 101:00	

審査請求 未請求 予備審査請求 有 (全 43 頁)

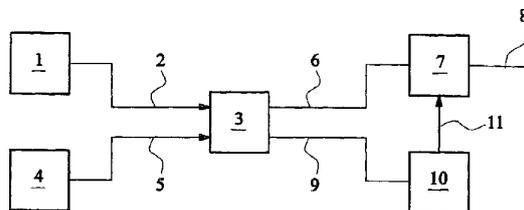
(21) 出願番号	特願2003-535459 (P2003-535459)	(71) 出願人	398038580 ヒューレット・パッカード・カンパニー HEWLETT-PACKARD COMPANY アメリカ合衆国カリフォルニア州パロアルト ハノーバー・ストリート 3000
(86) (22) 出願日	平成14年10月1日 (2002.10.1)	(74) 代理人	100099623 弁理士 奥山 尚一
(85) 翻訳文提出日	平成16年4月5日 (2004.4.5)	(74) 代理人	100096769 弁理士 有原 幸一
(86) 国際出願番号	PCT/GB2002/004458	(74) 代理人	100107319 弁理士 松島 鉄男
(87) 国際公開番号	W02003/032629		
(87) 国際公開日	平成15年4月17日 (2003.4.17)		
(31) 優先権主張番号	0123921.9		
(32) 優先日	平成13年10月5日 (2001.10.5)		
(33) 優先権主張国	英国 (GB)		
(81) 指定国	EP (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), JP, US		

最終頁に続く

(54) 【発明の名称】 自動写真撮影

(57) 【要約】

自動写真式カメラ装置または自動電子カメラ装置では、所定の人物からの音が優先的に受け取られる、または識別されて音声信号を提供し、それから該音声信号を使用して顕著性信号(11)(人物の状況に対する反応を示す)を生成し、カメラ1またはカメラ1から生成される画像信号(2、6)を制御する。音声は、たとえば人物の頭部または喉にある音伝達マイクロフォン(4)により、または必ずしも人物が装着する必要のない、特定性の低いマイクロフォンと併せて使用される声識別回路によって検出することができる。人物はカメラを装着しても、またはしていなくてもよい。制御はリアルタイムであっても、または後で画像信号を制御する(7)のために、音信号もしくは顕著性信号を画像信号とともに記録(3)してもよい。音声信号は、たとえば、発話(言葉認識および/または韻律)または非発話の音声、および/または非音声等の音を認識するために、波形解析によって分類することができ、信号の振幅も考慮に入れることができる。



【特許請求の範囲】**【請求項 1】**

画像信号を提供する電子カメラと、所定の人物からの音声を優先的に受け取りまたは識別し、これに応答して音声信号を生成することが可能な音検出手段と、前記画像信号と共に前記音声信号を記録する記録手段とを備える自動撮像装置。

【請求項 2】

前記音検出手段は、前記所定の人物の頭部領域または喉領域に装着されるように適合され、前記装着者の頭部から発せられる音声を優先的に検出し、これに応答して前記音声信号を生成する音伝達マイクロフォンを備える、請求項 1 に記載の自動撮像装置。

【請求項 3】

前記音声信号および画像信号を再生する手段と、前記再生音声信号に応答して顕著性信号を提供する音声手段と、前記顕著性信号に応答して、前記再生画像信号を制御する制御手段とをさらに備える、請求項 1 に記載の自動撮像装置。

【請求項 4】

カメラ、所定の人物からの音声を優先的に受け取り、または識別し、これに応答して音声信号を生成することが可能な音検出手段と、前記音声信号に応答して顕著性信号を提供する音声手段と、前記顕著性信号に応答して前記カメラを制御する制御手段とを備える自動撮像装置。

【請求項 5】

前記カメラは画像信号を提供する電子カメラである、請求項 4 に記載の自動撮像装置。

【請求項 6】

前記カメラは、画像を感光媒体に記録するように構成されている、請求項 4 に記載の自動撮像装置。

【請求項 7】

画像信号を提供する電子カメラと、所定の人物からの音声を優先的に受け取り、または識別し、これに応答して音声信号を生成することが可能な音検出手段と、前記音声信号に応答して顕著性信号を提供する音声手段と、前記顕著性信号に応答して前記画像信号を制御する制御手段とを備える自動撮像装置。

【請求項 8】

前記音検出手段は、前記所定の人物の頭部領域または喉領域に装着されるように適合され、前記装着者の頭部から発せられる音声を優先的に検出し、これに応答して前記音声信号を生成する音伝達マイクロフォンを備える、請求項 3 ~ 7 のいずれかに記載の自動撮像装置。

【請求項 9】

前記音声手段は、前記音声信号の波形を解析して、所定のカテゴリの集合の 1 つに入るものとして分類する音声解析手段を備える、請求項 3 ~ 8 のいずれかに記載の自動撮像装置。

【請求項 10】

前記音声解析手段は声音を認識するように構成されている、請求項 9 に記載の自動撮像装置。

【請求項 11】

前記音声解析手段は発話に関連する音声を認識するように構成されている、請求項 10 に記載の自動撮像装置。

【請求項 12】

前記音声解析手段は言葉を認識するように構成されている、請求項 11 に記載の自動撮像装置。

【請求項 13】

前記音声解析手段は韻律に関連する特徴を認識するように構成されている、請求項 11 または 12 に記載の自動撮像装置。

【請求項 14】

10

20

30

40

50

前記音声解析手段は非発話声音を認識するように構成されている、請求項 9 ~ 13 のいずれかに記載の自動撮像装置。

【請求項 15】

前記音声解析手段は非声音を認識するように構成されている、請求項 9 ~ 14 のいずれかに記載の自動撮像装置。

【請求項 16】

前記音声解析は一連の参照波形を含み、該一連の参照波形と照らし合わせて前記分類が行われる、請求項 9 ~ 15 のいずれかに記載の自動撮像装置。

【請求項 17】

前記音声解析は、新しい音声認識するようにトレーニング可能であるように構成されている、請求項 9 ~ 16 のいずれかに記載の自動撮像装置。 10

【請求項 18】

前記音声解析手段は、前記音声信号にカテゴリを割り当てるとき、全体の信号振幅またはエンベロープを考慮するように構成されている、請求項 9 ~ 17 のいずれかに記載の自動撮像装置。

【請求項 19】

空気を介して音声を受け取る従来のマイクロフォンをさらに備える、請求項 9 ~ 18 のいずれかに記載の自動撮像装置。

【請求項 20】

前記従来のマイクロフォンからの信号を記録する手段を備える、請求項 19 に記載の自動撮像装置。 20

【請求項 21】

前記音声解析手段は、分類に際し、前記音声信号を所定のカテゴリの集合の 1 つに入るものとして分類分類するとき、前記従来のマイクロフォンからの前記信号を考慮に入れるように構成されている、請求項 19 または請求項 20 に記載の自動撮像装置。

【請求項 22】

前記音声解析手段は、前記音声信号に割り当てられた前記カテゴリによって少なくとも部分的に決定される値を有する顕著性信号を提供するように構成されている、請求項 9 ~ 21 のいずれかに記載の自動撮像装置。

【請求項 23】

前記音声解析手段は、前記音声信号に割り当てられた前記カテゴリに従って二値の前記顕著性信号を提供する、請求項 22 に記載の自動撮像装置。 30

【請求項 24】

前記音声解析手段は、前記音声信号に割り当てられた前記カテゴリに従って複数の値のうちの 1 つを有する顕著性信号を提供する、請求項 23 に記載の自動撮像装置。

【請求項 25】

前記音声解析手段は、前記割り当てられたカテゴリの関数としての離散変数であり、前記音声信号の前記全体の信号振幅またはエンベロープの関数としての連続変数である顕著性信号を提供するように構成されている、請求項 22 に記載の自動撮像装置。

【請求項 26】

前記制御手段は、前記画像信号を選択的に伝送するゲートを備える、請求項 3 または請求項 7 に記載の、または請求項 3 または請求項 7 に直接もしくは間接的に従属する先行する請求項のいずれかに記載の自動撮像装置。 40

【請求項 27】

前記制御手段は、前記顕著性信号を他の制御信号と組み合わせて、前記ゲートを制御するゲート制御出力を生成する結合手段を備える、請求項 26 に記載の自動撮像装置。

【請求項 28】

前記ゲートは、前記二値の顕著性信号によって直接制御されるように構成されている、請求項 26 および請求項 23 に記載の自動撮像装置。

【請求項 29】

前記カメラには、少なくとも該カメラを前記所定の人物に固定する固定手段が設けられる、先行する請求項のいずれかに記載の自動撮像装置。

【請求項 30】

前記カメラには、少なくとも前記撮像光学系を前記所定の人物の頭部に固定する固定手段が設けられる、請求項 1 ~ 28 のいずれかに記載の自動撮像装置。

【請求項 31】

前記被制御画像信号を記録する記録手段を備える、請求項 7 に記載の、または請求項 7 に直接もしくは間接的に従属する請求項 8 ~ 30 のいずれかに記載の自動撮像装置。

【請求項 32】

前記記録手段には、該記録手段を前記所定の人物に固定する固定手段が設けられる、請求項 1 に記載の、または請求項 1 に従属する請求項のいずれか 1 項、または請求項 31 に記載の自動撮像装置。 10

【請求項 33】

前記記録手段は、音検出手段から離れたレコーダ位置に配置され、前記音声信号を前記音検出手段から前記レコーダ位置に無線伝送する手段を備える、請求項 1 に記載の、または請求項 1 に従属する請求項のいずれかに記載の自動撮像装置。

【請求項 34】

前記音声手段は、前記音検出手段から離れた音声手段位置に取り付けられ、前記音声信号を前記音声手段位置に無線伝送する手段を備える、請求項 3 ~ 7 のいずれか 1 項、または請求項 3 ~ 7 のいずれかに従属する請求項 8 ~ 30 のいずれかに記載の自動撮像装置。 20

【請求項 35】

前記制御手段は、前記音声手段から離れた制御位置に位置し、前記顕著性信号を前記音声手段から前記制御手段に無線伝送する手段を備える、請求項 3 ~ 7 のいずれか 1 項、または請求項 3 ~ 7 のいずれかに従属する請求項 8 ~ 30 のいずれかに記載の自動撮像装置。

【請求項 36】

前記カメラは、前記所定の人物上以外のカメラ位置に取り付けられる、請求項 33 ~ 35 のいずれかに記載の自動撮像装置。

【請求項 37】

電子カメラを操作して画像信号を生成させる、操作するステップと、所定の人物からの音声を優先的に受け取り、または識別し、これに回答して音声信号を生成するステップと、前記音声信号を前記画像信号とともに記録するステップと、を含む撮像方法。 30

【請求項 38】

前記選択的に識別するステップは、前記所定の人物の頭部から発せられる音声を優先的に検出する、前記所定の人物の頭部領域または喉領域に合った音伝達マイクロフォンを使用して行われる、請求項 37 に記載の撮像方法。

【請求項 39】

前記音声信号および前記画像信号を再生するステップと、前記再生音声信号に回答して顕著性信号を生成するステップと、前記顕著性信号に回答して前記再生画像信号を制御するステップと、をさらに含む、請求項 37 に記載の撮像方法。

【請求項 40】

カメラを操作して画像信号を生成させる、操作するステップと、所定の人物からの音声を優先的に受け取り、または識別し、これに回答して音声信号を生成するステップと、前記音声信号に回答して顕著性信号を生成するステップと、前記顕著性信号に回答して前記カメラを制御するステップとを含む撮像方法。 40

【請求項 41】

前記カメラは画像信号を提供する電子カメラである、請求項 40 に記載の撮像方法。

【請求項 42】

カメラを操作して画像信号を生成させる、操作するステップと、所定の人物からの音声を優先的に受け取り、または識別し、これに回答して音声信号を生成するステップと、前記音声信号に回答して顕著性信号を生成するステップと、前記顕著性信号に回答して前記画 50

像信号を制御するステップと、を含む撮像方法。

【請求項 4 3】

前記優先的に受け取り、または識別するステップは、前記所定の人物の頭部から発せられる音声を優先的に検出する、前記所定の人物の頭部領域または喉領域に合った音伝達マイククロフォンを使用して行われる、請求項 3 9 ~ 4 2 のいずれかに記載の撮像方法。

【請求項 4 4】

前記顕著性信号を生成するステップは、前記音声信号の波形を解析して、所定のカテゴリの集合の 1 つに入るものとして分類することを含む、請求項 3 9 ~ 4 3 のいずれかに記載の撮像方法。

【請求項 4 5】

添付図面を参照して実質的に上に述べた自動撮像装置。

【請求項 4 6】

添付図面を参照して実質的に上に述べた撮像方法。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

写真撮影の自動化は常に関心の対象であり、過去数十年にわたって高い関心を集めている。自動露出制御およびオートフォーカスは、今では大半のカメラで標準の機能であり、多くのカメラは他の電子的に制御される機能、または電氣的機能を有している。1 つ以上の感光アレイの形態の電子イメージセンサを備えたビデオカメラおよびスチルカメラは、ますます一般での入手が可能になりつつあり、画像を安価に記録する手段、編集を選択して、受け入れられると考えられる画像のみをプリントする手段、PC および付属機器のみを使用して家でこれを行う手段、および電子メールを使用して画像を送受信する手段を含め、以前は利用できなかった機能を提供することから人気がある。

【0 0 0 2】

しかし、一般に、買ってきただけのカメラでは、写真のフレーミング（方向およびズーム）ならびにタイミングについての決定を含めて、操作者が撮影する実際の写真の構成を決定する必要がある。電子カメラの出現により、複数の写真を連続して、またはビデオクリップさえも記録し、そこからしかるべき画像を選択することができ、また、この後にある程度の編集を行うことができるが、操作者はやはり、カメラをおおよそ正しい方向に向けておおよそ正しいときに作動ボタンを押下する必要がある。

【0 0 0 3】

スチルまたはビデオ電子カメラは両方とも、一連の画像信号（「画像シーケンス」）を提供可能なセンサを有するため、本明細書では広く「電子カメラ」と呼ぶことにし、それぞれの出力を、得られるのが静止画像であるか、それとも画像シーケンスであるかに関わりなく、また画像信号がカメラに記録されるのか、それとも他のどこかに記録されるのかに関わりなく、「画像信号」と呼ぶことにする。写真または画像というときは、静止画像および画像シーケンスの両方を対象として含むものとして読むべきである。「カメラ」と限定することなく述べるときは、電子カメラあるいは感光記録媒体を使用するカメラ、たとえば、静止露出用のフィルム式カメラまたはシネカメラ（「写真」カメラ）に関連するものとして読むべきである。

【0 0 0 4】

オートフォーカスおよび露出等のカメラの制御は、写真撮影の技術的側面により関連する。これらは、フレーミングおよびタイミングを含む写真の構成にはほとんど関係ないか、またはまったく関係ないが、現在、この側面に関心が集まり、および意識のある操作者による入力がさらに低減されているか、または完全になくなった写真撮影機能または写真選択機能を有するインテリジェントなスチルカメラおよびビデオカメラの提供に対して関心が集まっている。

【0 0 0 5】

考えられるある方策は、連続して動作するカメラからの画像信号全体を格納する、または

10

20

30

40

50

シネカメラからのシーケンス、すなわち画像シーケンスの全体を格納し、そこから適した単一画像または少なくなった画像シーケンスを後の時間期間に選択するというものである。これにより、カメラ操作者に対する負担が軽減されるか、またはカメラ操作者に対する必要性が全く回避されるが、この手法では非常に大量の画像信号の記憶容量またはフィルム格納場所が必要であるだけでなく、操作する人が、後の選択プロセス中に、格納されている画像シーケンスを調べる必要があり、これもまた非常に時間がかかり、ひいては高価である。したがって、格納前に少なくともある程度の画像を選択させるか、あるいは画像シーケンス全体が格納される場合、画像シーケンスとともに、シーケンスの最も関心のある部分が見つけれられる可能性が高いところを示すさらなる信号を記録することが非常に望ましい。

10

【0006】

ある程度の写真の構成が行われる既知の種類 of 自動カメラ装置は、多様なものを包含している。一例としては、写真撮影状況が厳密に予め決定され、かなり限定された機器の形態であり、たとえば、本質的に固定位置にある席に着いた顧客が機械にお金を入れた後、カメラが固定されたタイミングで動作するパスポート写真撮影用のブースがある。

【0007】

所定の条件または条件の集合に応答して、たとえば、1つ以上のセンサの出力に対して働くアルゴリズム的計算の結果に応答して、自動的に写真撮影が始まる装置では、より自由が許される。しかし、この種類の装置の1つの問題は、写真撮影を可能にするために必要な条件の集合を定義することにある。

20

【0008】

ある特定の状況では、規定された条件が物理的に予め決定されている。警報の入力に応答して動作を開始するセキュリティカメラ、および運転の速すぎる車両に応答する交通速度カメラは、このカテゴリに入る。また、来訪者が乗っている車が、赤外線ビームを遮断することなどによってスイッチを作動させたときに起動される適宜に配置されている固定カメラを起動させて、テーマパークに車で来た来訪者の写真を撮影することも知られている。このようなシステムは、たとえば、国際特許出願番号WO98/10358号(Goldberg)、欧州特許出願第0953935号(Eastman Kodak)、米国特許第5,655,053号(Renie)、および米国特許第5,576,838号(Renie)に開示されている。

【0009】

他のシステムによっては、撮像すべき特徴に関連する特徴の認識または検出によって幾分許容範囲が広いものもある。したがって、米国特許第5,844,599号(Hildin)には、会議テーブルの周囲の一群の人々と使用されるビデオカメラシステムが開示されており、各人がマイクロフォンを装着し、カメラは、そのときに話している人に向けられ、その人の動きをトラッキングするように構成される。これは、画像の記録よりも閲覧を意図したシステムである。欧州特許出願第0660131号(Osen)には、全地球測位システムを採用して、航空機等の特徴を、カメラを自動的にそこに向けることができるようにトラッキングするシステムが記載されている。

30

【0010】

この点に関して、撮像中の特徴を識別し、かつ/またはカメラを起動するタグがますます利用されるようになってきている。これは、上に述べた国際特許出願番号WO98/10358号(Goldberg)、および米国特許第5,694,514号(Evans)で行われ、異なる特徴の写真の容易に識別することが可能である。国際特許出願番号WO98/10358号(Goldberg)には、タグを使用してカメラを起動し、タグを装着している人に画像を関連付けるデータを提供することを含む複数のカメラシステムが記載されており、一変形では、顔認識を用いて特徴を識別し、最初にシステムに与えられる画像および他の関連データから作業を着手する。

40

【0011】

しかし、いつでも、また関心があると考えられるあらゆるものの写真を撮影しようとするどんなところでもハンドヘルドカメラを用いることができる個人のカメラユーザと比較し

50

て、上に述べた従来技術のすべてでは、写真撮影の様式に関してかなりの制限が残っている。最もよく知られている自動カメラ装置は、固定カメラを採用するか、かつ/またはカメラが起動される状況に対して厳密な制約を有するか(たとえば制限された条件または条件の集合に関連するタグ、条件、またはアラーム入力への応答すること)のいずれかである。

【0012】

テーマパーク等の商業的に設置されたものでは、カメラは本質的に固定され、特定の状況において画像を取り込むように配置される。この特定の状況とは、たとえば、格納される画像の少なくともいくつかは満足に行くように構成され、購入者になりうる人にとって興味のあるものとして心を引くことが証明されることになるものと期待される場合である。

10

【0013】

可搬式であり、手動で操作されるカメラの場合のように、人物の位置に関わりなく、異なる画像を自動的に取り込むことができるカメラを提供可能なことが望ましい。

【0014】

したがって、自動カメラ装置のさらに望ましい特徴は、見られている画像は所定の種類に関連しないかもしれないが、それでもある人物、たとえばカメラのユーザにとって興味のあるものでありうることを示す信号(「顕著性(saliency)」信号)の提供であると考えることができる。この信号は、選択的なカメラ操作の起動、または選択的なリアルタイムでの画像の格納に使用することができ、または、後の画像選択プロセスにおいて連続して格納される画像信号とともに使用することができ、もし信号が3つ以上の値(すなわち関心のある特徴に画像が関連するという相対的な尤度を示す)を有する場合には、これら技法を組み合わせたものが採用される。すなわち、関心のある可能性のあるすべての信号を格納し、他の信号を拒絶し、格納されている信号から後に選択し、初期の優先がより関心のあるものと示される信号の部分に与えられる。このような特徴は、本発明に関して意図される。

20

【0015】

人物から発せられる音は、その人物がいる環境に対する人物の反応を示すことができるということが分かっている。このような反応は、特定の出来事に対するものであったり、または環境の全体的な雰囲気に対するものであったりする。反応は、たとえば、見ている、または可視のシーンまたは出来事、別の人物との対話、突然の音(ボタン等)であるか、または連続した音(音楽等)であるかに関わらず音、および出来事に参加しているとき、または遊園地の乗物等の娯楽を楽しんでいるときに受ける刺激等の他の刺激に対してのものであることができる。反応がどうであれ生じると、人物が見ているシーンに、その反応の原因を表すか、または後日、少なくともその反応の原因を呼び起こすものであることができる。したがって、本発明の実施形態では、人物から発生された音が、顕著性信号の生成に採用される。本発明は、可搬性である、すなわち人物が携帯するカメラに対して特に有用であるがそれに限定されず、動くことのできる無生物物体に取り付けられるカメラ、または固定位置に取り付けられるカメラ、または固定軌道を辿るカメラと使用することが可能である。

30

【0016】

第1の態様では、本発明は、画像信号を提供する電子カメラと、所定の人物からの音を優先的に受け取り、または識別し、これに応答して音声信号を生成することが可能な音検出手段と、画像信号とともに音声信号を記録する記録手段とを備える撮像装置を提供する。この第1の態様では、本発明は、音声信号および画像信号を再生する手段と、再生音声信号に応答して顕著性信号を提供する音声手段と、顕著性信号に応答して再生画像信号を制御する制御手段とをさらに備える。

40

【0017】

第2の態様では、本発明は、カメラと、カメラと、所定の人物からの音を優先的に受け取り、または識別し、これに応答して音声信号を生成することが可能な音検出手段と、音声信号に応答して顕著性信号を提供する音声手段と、顕著性信号に応答してカメラを制御す

50

る制御手段とを備える撮像装置を提供する。

【0018】

第2の態様では、好ましくは、カメラは画像信号を提供する電子カメラであるが、従来のスチルカメラまたはシネフィルムカメラであってもよい。カメラの制御は、写真撮影を起動することとすることができる。

【0019】

第3の態様では、本発明は、画像信号を提供する電子カメラ、所定の人物からの音を優先的に受け取り、または識別し、これに回答して音声信号を生成することが可能な音検出手段、音声信号に回答して顕著性信号を提供する音声手段、および顕著性信号に回答して画像信号を制御する制御手段を提供する。

10

【0020】

本発明は、第1～第3の態様による装置を使用して撮像する方法に拡張される。

【0021】

後に特に述べるように、好ましい音検出手段は、装着者（「所定の人物」）の頭部に合うように適合され、装着者の頭部および/または喉から発せられた音を優先的に検出し、これに回答して音声信号を生成する音伝達マイクロフォンである。しかし、所定の人物からの音を優先的に受け取る、または識別する他の技法も既知であり、音伝達マイクロフォンを使用する代わりにこれらを使用することも本発明の範囲内である。たとえば、これら技法は、2人以上の人物からの音を受け取ることができるマイクロフォンに対して動作する声認識ソフトウェアを使用することを含み、このようなソフトウェアは、特に特定のユーザからの、発話を認識するようにトレーニングすることができる。それにも関わらず、多くの場合、音伝達マイクロフォンを使用することで、意図する機能を実行する特に単純かつ効果的な方法を提供することができる。

20

【0022】

本発明の3つの態様すべてに共通の装着式音伝達マイクロフォンは既知であり、通常、装着者の体、頭部および耳に装着されるマイクロフォンの場合は頭蓋骨を通して伝達される音の振動を検出する振動変換器を備える。これらマイクロフォンは、空気を通して伝達される音を拾うマイクロフォンと区別される。

【0023】

耳装着式マイクロフォンの典型例は、国際特許出願番号WO 86 / 07157 (ACR Electronics)、米国特許第4,588,867号 (Konomi)、および米国特許第5,692,059号 (Kruger) に記載されている。他の頭蓋装着式マイクロフォンによっては、頭頂部に固定されるものがある。これらはすべて、通常は骨伝導を介して頭部を通して直接伝達される音を拾う振動変換器を備えている。別の典型的な形態の頭部装着式マイクロフォンは、喉に物理的に接して装着される。空気を通して伝達される音もまた、こういったマイクロフォンによって感知することができる。このようなマイクロフォンを動作させる回路は、たとえば、国際特許出願番号WO 00 / 21194 (Resound Corporation) および欧州特許出願第0984660 (Nippon Telephone and Telegraph Corporation) に開示されている。

30

【0024】

音信号によって起動するカメラは既知である。米国特許第5,014,079号、および第5,107,288号 (両方ともKonica Corporation)、ならびに米国特許第5,162,832号 (Konishoroku Photo Ind.) に例が開示されているが、これらは概して、特定の1つの音特徴、たとえば、これら特定の場合作の強度エンベロープ (すなわち音圧) にのみ応答するように構成されており、マイクロフォンはカメラに配置されており、音伝達マイクロフォンではない。

40

【0025】

また、米国特許第5,844,599号 (Lucent Technologies Inc.) から既知なのは、各参加者に、その人物が話しているときにタグとして働くマイクロフォンが提供される、会議で使用する自動カメラシステムであり、このシステムは、カメラが自動的に、話者また

50

は最も大きな声の話者の方を向くように構成される。このシステムは個人的なマイクロフォンを含むが、特に音伝達マイクロフォンではなく、利用される音パラメータは音の強さのみである。

【0026】

この従来技術はいずれも、カメラの動作の制御またはカメラから生成される画像信号の制御に、頭部または喉に装着された音伝達マイクロフォンからの音声信号を使用することを対象としていない。

【0027】

音を分類するシステムも存在し、上に述べた音声手段は、音声信号の波形を解析して、所定のカテゴリの集合の1つに入るものとして分類する音声解析手段を含むことができる。

10

【0028】

米国特許第5,712,953号(Langs)は、音声信号を音楽または非音楽として分類するシステムを開示しており、他のこのような従来技術を参照している。おそらく、本発明により関係するものとして、米国特許第5,918,223号(Muscle Fish)は、所与の音信号と比較することによって音声信号をより詳細に分類するシステムを開示している。このシステムは、マルチメディアデータベースアプリケーション、および所与の音と同様に聞こえる音声データファイルを見つけるインターネットサーチエンジンで有用であると言われている。所与の一例は、サウンドトラック中の拍手および声援の識別である。ユーザが、「鳥の声」または「ロックミュージック」等の音の分類を作成することができる。

【0029】

本発明の装置の基本的な形態では、音声解析手段は、所定の人物が発した、笑い声等の特定の1つのみの形態の音を認識するように適合することができ、このような場合、最も基本的なカテゴリの集合は「笑い声」および「他」からなるであろう。

20

【0030】

しかし、他よりも困難なものもあるが、所定の人物が発した複数の音を分類し、そうして識別可能なカテゴリおよび/またはカテゴリが関連する出来事の範囲を広げることが可能である。所定の人物の頭部からの音を、別個に考える必要がある非発話および発話に分けることができる。

【0031】

典型的な非発話音は、笑い声、鳴き声、鼻歌、咳、およびくしゃみ等の声音と、鼻で吸い込む音、引っ掻き音、食べているときの音、歯を摺り合わせる音(歯ぎしり)、およびげっぷ等の非声音を含む。

30

【0032】

発話自体は、言葉と、言葉が話された様式(韻律)とに分けることができる。言葉認識技法は、多くの言葉を正確に認識するに十分に進歩している。韻律は、声の強度パターン(上がり調子の声か、それとも下がり調子の声か、または一続きの文章または詩の朗読、問い掛け、もしくは陳述に関連する変動パターン、ためらい、言葉が発せられた様式、すなわち、音節に関連する異なる強勢、声の周波数、話す速さ、発話の部分部分の持続時間のよう特徴を含む。これは、離散した抽象的な発話単位を識別する音韻の側面と、このような単位に関連する連続変化する音響情報に関連する音声学的な側面とからなり、所定の

40

【0033】

歌声等、音によっては、発話と非発話の両方の関連に入ると考えられるものもある。

【0034】

本発明では、音声解析手段によって行われる分類はさらに、全体の信号(エンベロープ)の振幅の関数であることができるが、これは、発話が含まれるいずれの場合でも韻律の一部であると考えることができる。

【0035】

したがって、音声解析手段は、声音を認識し、その音に発話に関連しているか、それとも非発話に関連しているかを認識するように構成することができる。発話が関わっている場

50

合、音声解析手段は、言葉を認識し、かつ/または韻律に関連する特徴を認識するように構成することができる。しかし、代替または追加として、音声解析手段は、非発話声音を認識し、かつ/または非声音を認識するように構成してもよい。

【0036】

加えて、音声解析手段は、カテゴリを音声信号に割り当てるときに全体の信号振幅またはエンベロープを考慮に入れるように構成することができる。たとえば、大きな笑い声と静かな笑い声とを区別するように構成することが可能である。

【0037】

音声解析には、所定の一連の参照波形を提供することができ、この参照波形と照らし合わせて分類を行う。代替または追加として、また好ましくは、新しい音を認識して、所定の人物の特定の癖に適合することができるようにトレーニング可能なように構成することができる。

10

【0038】

言葉認識は、自発的および非自発的なカメラまたは画像信号制御に有用であることができる。「こんにちは」、「さようなら」、「うわっ」、および「続けて」等の言葉および表現は、人々を相手にしているとき、またはイベントに参加しているときに写真撮影を自発的に開始する際に有用であることができ、一方で、「それを撮影」等のコマンドは、写真を撮影する故意のコマンドとして認識することができる。得意げに話される「はい」、または疑念がある場合などに話される「いいえ」等の言葉は、韻律と組み合わせられる言葉認識の例である。笑い声および鼻歌は、写真の撮影または選択を開始することができる非発話声音の2つの例として解釈することができる。大きな笑い声は、条件なしでこのような開始を行うことができるが、柔らかな笑い声は、これが行われるか否かを判断するために、さらなる条件信号と組み合わせる必要がありうる。

20

【0039】

顕著性信号は、こうして行われる分類に従って生成することができる。最も単純かつ可能な限り最も実用的な形態では、顕著性信号は二値であり、したがって、マイクロフォンからの音に応答して、「はい」または「いいえ」信号が常に発行される。これは、単独で利用することもでき、または制御信号として使用される他の制御信号と論理的に組み合わせられて利用することもできる。たとえば、制御信号は、たとえばカメラに写真を撮影するように言うことによって、もしくは撮像シーケンスを開始するように言うことによって、または画像信号を伝送、表示、または記録するようにゲートを動作させることによって、カメラまたは画像信号を制御する際に使用することができる。

30

【0040】

代替として、顕著性信号は、ある範囲の離散値または連続値を有することができる。これは、特定の値の顕著性信号に関連する特定の音を識別するために、あるいは異なる音に割り当てられた所定順序の値によって定められる最も関心のある音をサーチするために、完全な画像シーケンス全体を通してスキャンする人にとって有用でありうる。また、このような形態では、たとえば、所定のアルゴリズムに従って他の信号と数値的に組み合わせて、おそらく、結果得られる信号をしきい値処理した後、前の段落で述べた種類の制御信号を生成することができる。

40

【0041】

数値的な顕著性信号を導出するには、認識された異なる音（またこれは、認識された異なる音であるものとして異なる強さを有する同様の音を含むことができる）に異なる値を割り当てることができる。これで十分であるか、または強さがまだ考慮に入られていない場合、割り当てられた値を音の強さに従って連続して変更することができる。

【0042】

本発明の好ましい形態では、カメラは所定の人物、好ましくは音伝達マイクロフォン装着者によって装着され、好ましくは、装着の方の前方に向けられるように装着される。カメラは胸部に、たとえばバッジとして装着してもよく、カメラ、または少なくともカメラの撮像光学系を人物の頭部、たとえば、1つの眼鏡のフレーム、また帽子等のかぶりものに

50

設置する固定手段がカメラに設けられることが特に好ましい。このようにして、カメラの方向は、装着者の注視方向をより正確に辿る。適切であれば、カメラの視野を、おそらく目の動く全範囲をカバーするように、人である装着者の視野よりも大幅に広くすることができる。これは、たとえば固体状態カメラ中のセンサレイの利用可能なピクセル数が、通常の人々の視野を近似するビュー用に解像度を維持するように対応して増大する場合に得に有利である。

【0043】

本発明の第1の態様における真に独立した可搬性のため、たとえば、再生画像信号を制御する顕著性信号を後に生成するために、または見ている人の関心がありうる再生画像信号の部分を示すために、所定の人物に、画像信号および音声信号を両方とも記録するレコーダも設けられる。

10

【0044】

本発明の第2および第3の態様における真に独立した可搬性のため、所定の人物に、顕著性信号を提供する音声手段も設けることができる。次にこれは、先の段落で述べたように使用するために、ディスプレイまたはレコーダへゲーティングすることなどによって画像信号を制御するために、またはカメラを制御または起動するために、完全な画像信号とともに記録することができる。

【0045】

しかし、音声信号、顕著性信号、および画像信号はいずれも、適切であれば離れた場所に伝送することができる。音声信号または顕著性信号の伝送が適切である一機会は、カメラが所定の人物に装着されていないときである。たとえば、装置は、カメラが所定のシーンを見るように固定されている設備の一部であることができ、それから人物が撮影者として働き、自発的に（言葉でのコマンド）または非自発的に（非発話音）カメラの動作、画像信号のゲーティングを制御するか、または少なくとも、記録された信号の顕著性を示す。これは、誕生日パーティ等の特別な場合に使用するものでありうる。カメラは、固定する必要はなく、たとえば、所定の人物をトラッキングしている遊園地にある乗物場内の所定のトラックを辿ることもできることに留意されたい。読み手には、本発明の他の適した構成および用法が思い浮かぶことを確信する。

20

【0046】

本発明の好ましい実施形態の必須部分は音伝達マイクロフォンの使用であるが、1つまたは複数のさらなるマイクロフォンを設けることもでき、また、少なくとも1つのこのようなさらなるマイクロフォンは、空気を介して音を受け取る「従来の」マイクロフォンであってもよいことに留意されたい。

30

【0047】

「従来の」マイクロフォンからの信号は、たとえば、後日再生するために、画像信号および/または音伝達マイクロフォン信号とともに記録することができる。この点において、聞く目的では「従来の」マイクロフォンからの信号を再生するほうがより普通であり、音伝達マイクロフォンからの信号は、この目的では現実のものではない音を生成することを念頭に置かれたい。

【0048】

代替または追加として、この信号を、音伝達マイクロフォンからの信号と組み合わせてもよい。たとえば、音伝達マイクロフォンを装着している人物が生成する音の大半を、「従来の」マイクロフォンが生成する信号から差し引いて、音伝達マイクロフォン装着者周囲の音と考えられるものを残すことが可能でありうる。またこれにより、周囲音のパターンおよび伝達マイクロフォンによって検出されたパターンを解析する、または比較することなどによって、伝達マイクロフォンの装着者が別の集団との会話に関わっていたか否かについて推定することができる。これは、顕著性信号を決定する特徴であることができる。すなわち、マイクロフォンの組み合わせの出力を解析することによって顕著性信号を導出することができる。このような組み合わせに「従来の」マイクロフォンが含まれてもよい。

40

【0049】

50

カメラは、静止写真および/または画像シーケンスを取り込むことが可能でありうる。先の段落で言及したように、「従来の」マイクロフォンを備えてもよい。

【0050】

読み手は、大半の場合、それ自体が既知のフリー工技法を利用することによって音声解析をより効果的かつ効率的に行うことができることを理解するであろう。

【0051】

本発明のさらなる特徴および利点は、読み手が参照する添付の特許請求の範囲を考慮すること、および図1～4が本発明の第1～第4の実施形態を概略的な形で示し、図5が音声解析を概要略図の形態で例示する添付図面を参照して行われる以下の本発明の好ましい実施形態のより詳細な説明を読むことにより、明確になるであろう。

10

【0052】

しかるべき場合、同様の番号が各図面中の同様の要素を示す。

【0053】

図1では、頭部に装着された電子カメラ1が画像信号2を生成し、マルチチャンネルレコーダ3に格納される。耳装着式音伝達マイクロフォン4が音声信号5を生成し、画像信号と同期してレコーダ3に記録される。その後、再生中に、再生音声信号9が音分類回路10に渡され、音分類回路10が、音波形の形状が格納されている複数の波形の形状のいずれか1つに近似するか否かによって二値(顕著性)信号11を生成する。信号11を使用して、再生画像信号6を選択的に伝送(8)するようにゲート7を動作させる。一変形では、特に注目されるべきことを示すため、信号11を使用して、再生され表示される画像信号6の人である閲覧者に対して合図を送る。

20

【0054】

図2では、頭部装着式音伝達マイクロフォン4からの信号5が、回路10において分類され、一方向における二値(顕著性)出力11の変化を利用して、単一の写真を撮影するように、または画像シーケンス(ビデオまたはシネクリップ)を開始するように、写真または電子カメラ1を動作させる。後者の場合、クリップは、信号11の状態が他方の方向に変化するまで続けられる。また、電子カメラ1の場合、画像信号2が12において記録または表示される。

【0055】

図3では、喉装着式音伝達マイクロフォン4の出力5が、回路10において分類されて、画像信号2をレコーダまたはディスプレイ12に選択的に伝送するゲート7を制御する顕著性信号13を生成する。信号13は、二値信号であってもよく、また、ゲート7に直接接続してもよいが、図示のように、ゲート7を制御する二値信号15を生成する回路14内の1つまたは複数の他の信号16と結合される。

30

【0056】

この回路の第1の形態では、信号13および信号(複数可)16は二値信号であり、回路14において論理的に結合される。たとえば、信号13は信号16と論理積演算され、写真構成(画像信号2の空間的な分割によって得られる)が許容できるものである事を示し、その結果を信号16と論理和演算し、カメラにいずれの場合でも写真を撮影するように指示する。後者の信号は、実際には、「それを撮影(take that)」等のコマンドの検出による回路10からの信号であることができるが、必ずしもそうである必要はない。

40

【0057】

この回路の第2の形態では、回路10は、検出された音の種類に従って音声信号5に値(顕著性信号)を割り当て、これらの値がアルゴリズム回路14において、所定のアルゴリズムに従って他のソースからの数値信号16と結合され、出力がしきい値処理されて、二値信号15を提供する。この場合、信号16の値に関わりなく信号15の値を決定するように、「それを撮影」等のオーバーライドコマンドが極値に割り当てられることになる。

【0058】

図4では、耳に装着された音伝達マイクロフォン4からの信号5が、無線接続18、19を介して、たとえば、RF、赤外線、または誘導を介してマルチチャンネルレコーダ3に伝

50

送され、マルチチャンネルレコーダ 3 はまた、固定カメラ 17 から画像信号出力 2 を、少なくとも 1 つの他のソースから制御信号 (複数可) 16 を受信する。その後、信号は画像信号 6 および制御信号 20 として再生され、制御信号 20 は回路 14 によって処理され、図 3 のように画像信号ゲート 7 の制御に使用される。

【0059】

図 4 の一変形では、カメラは遊園地乗物場での乗物に固定され、乗車間にマイクロフォン装着者を観察する。信号 16 の 1 つは、乗車が進行しているときの時間期間およびその前後の短時間期間を含む時間期間に提供されるイネーブル信号である。

【0060】

図 5 は、顕著性信号 27 を提供する音声解析装置を概略的形態で示す。音声信号 5 は、波形形状解析回路 21 およびエンベロープ振幅解析回路 25 で受け取られる。

10

【0061】

回路 21 は、波形の種類を決定し、その種類が声 - 発話、声 - 非発話、非声、または決定不能のそれぞれであると決定されたことに従って 4 つの信号 29 ~ 32 のセットのうちの 1 つを生成する。信号 29 ~ 31 はそれぞれ、各回路 22 ~ 24 を使用可能にして、格納されている各波形セットそれぞれと波形形状が一致する程度を決定するために使用される。次に、回路 22 ~ 24 のうちの動作している回路は、十分な対応がある場合、入力波形が最も密接に対応する格納波形を一意に識別するラベル信号 33 の生成に進み、その他の場合、対応がないことを示すラベル信号を提供し、これは、回路 21 からの信号 32 に等しい。

20

【0062】

信号 33 ~ 35 は、回路 25 からのデジタル化された音量 / 振幅信号 36 と結合され、これとともに使用されて、二次元ルックアップテーブルをアドレス指定する。テーブルからの二値出力 27 の場合、テーブルコンテンツはすべて 0 と 1 のいずれかであるが、数値顕著性出力の場合、テーブルはある範囲の数値を含みうる。

【0063】

全体波形の指示がまったくない状態で、エンベロープ振幅の決定が問題である場合、回路 21 および 25 が結合されて、破線 28 で示すように、波形形状が決定された後にエンベロープ振幅を決定することに役立つ。代替または追加として、出力 33 ~ 35 を回路 25 に結合してもよい。

30

【0064】

さらに、韻律が重要な場合、信号 5 とともに回路 22 の出力 33 を、テーブル 26 に入力される代替の数値出力を提供する韻律決定回路に結合することができ、この場合の数値出力は、言葉および関連する韻律を示す。

【国際公開パンフレット】

(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)

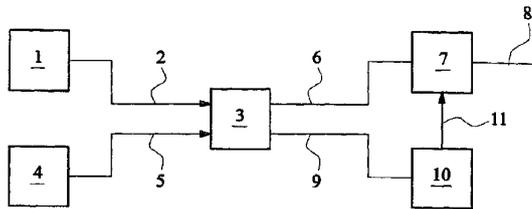
PCT

(10) International Publication Number
WO 03/032629 A1

- (51) International Patent Classification: 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: 5 October 2001 (05.10.2001) GB 0123921.9
 - (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

WO 03/032629

PCT/GB02/04458

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
5 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
10 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
15 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
20 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
25 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

WO 03/032629

PCT/GB02/04458

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.

5 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
10 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.

15 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
20 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
25 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

WO 03/032629

PCT/GB02/04458

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.

WO 03/032629

PCT/GB02/04458

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.

- 5 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.

10 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
15 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.

20 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
25 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
30 in the production of a saliency signal. The invention finds particular use with cameras

WO 03/032629

PCT/GB02/04458

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

WO 03/032629

PCT/GB02/04458

“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

WO 03/032629

PCT/GB02/04458

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
5 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.

10 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
15 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
20 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.

25 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".

WO 03/032629

PCT/GB02/04458

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
5 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
10 (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
15 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.

20 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.

25 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

WO 03/032629

PCT/GB02/04458

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
5 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.

10 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
15 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
20 effected.

A saliency signal may be developed according to the categorisation thus performed. In its simplest, and possibly 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
25 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.

WO 03/032629

PCT/GB02/04458

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
5 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
10 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,
15 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
20 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
25 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
30 interest to the viewer.

WO 03/032629

PCT/GB02/04458

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

WO 03/032629

PCT/GB02/04458

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
5 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
10 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
15 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.

20 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
25 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

WO 03/032629

PCT/GB02/04458

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

5 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.

10 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.

15 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
20 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
25 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.

WO 03/032629

PCT/GB02/04458

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

WO 03/032629

PCT/GB02/04458

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
5 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.

10

WO 03/032629

PCT/GB02/04458

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.
5
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.
10
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.
- 15 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.
- 20 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.
- 25 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

WO 03/032629

PCT/GB02/04458

- 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.

WO 03/032629

PCT/GB02/04458

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 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.

WO 03/032629

PCT/GB02/04458

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
5 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
10 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.
- 15 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,
20 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.
- 25 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

WO 03/032629

PCT/GB02/04458

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
5 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
10 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
15 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
20 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
25 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.

WO 03/032629

PCT/GB02/04458

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.
- 5
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.
- 10
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.
- 15
46. A method of imaging substantially as hereinbefore described with reference to the accompanying drawings.

WO 03/032629

PCT/GB02/04458

-1/2-

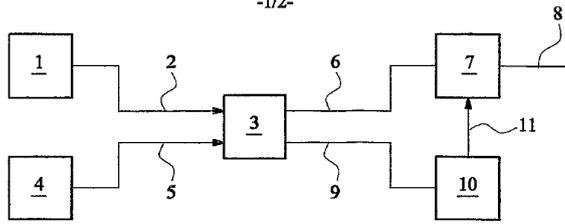


FIG. 1

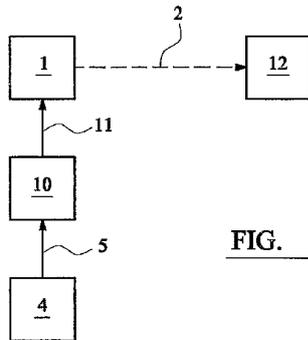


FIG. 2

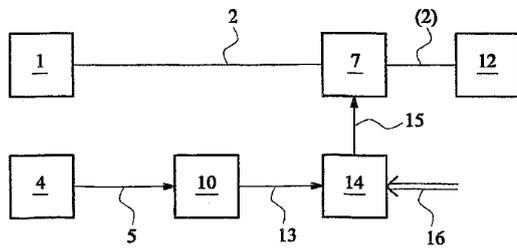


FIG. 3

WO 03/032629

PCT/GB02/04458

-2/2-

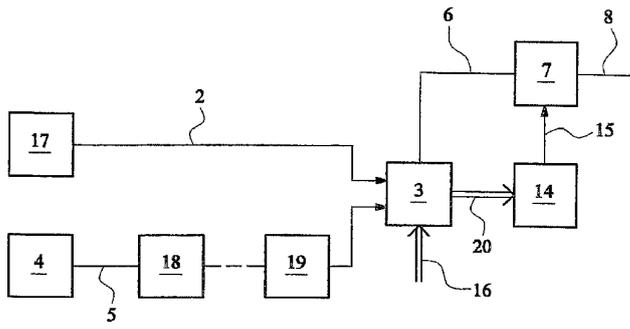


FIG. 4

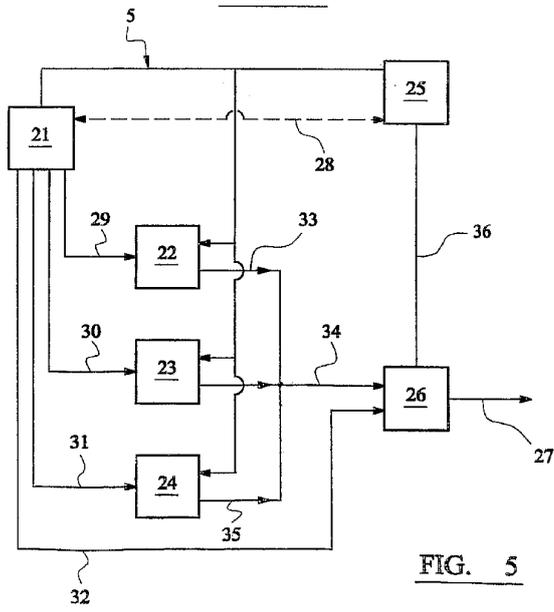


FIG. 5

【 国際調査報告 】

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 610L		
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 BRIAN) 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 25,27,35
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 --- -/--	
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		
<input checked="" type="checkbox"/> 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. *Z* 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 European Patent Office, P. B. 2918 Patentsstr. 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 apo nl, Fac. (+31-70) 340-3016		Authorized officer Didierlaurent, P

Form PCT/ISA/210 (second sheet) (July 1999)

INTERNATIONAL SEARCH REPORT		PCT/GB 02/04458
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

Form PCT/ISA210 (continuation of second sheet) (July 1992)

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

Form PCT/GB/R10 (patent family annex) (July 1998)

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. <input type="checkbox"/>	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2. <input type="checkbox"/>	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. <input type="checkbox"/>	Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(e).
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. <input checked="" type="checkbox"/>	As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. <input type="checkbox"/>	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. <input type="checkbox"/>	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. <input type="checkbox"/>	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	<input checked="" type="checkbox"/> The additional search fees were accompanied by the applicant's protest. <input type="checkbox"/> 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

フロントページの続き

(72)発明者 グロヴナー, デイヴィッド・アーサー
イギリス国, ビーエス36 2エイディー ブリストル, フランプトン・コットレル, ビューフォ
ート・ロード 18

Fターム(参考) 5C122 DA03 DA04 EA42 EA69 FJ01 FJ04 FJ15 GA31
5D015 KK02