(19) **Bundesrepublik Deutschland**
Deutsches Patent- und Markenamt

(10) **DE 101 56 954 B4** 2004.12.23

(12) # Patentschrift

(21) Aktenzeichen: **101 56 954.8**
(22) Anmeldetag: **20.11.2001**
(43) Offenlegungstag: **18.06.2003**
(45) Veröffentlichungstag
der Patenterteilung: **23.12.2004**

(51) Int Cl.[7]: **G10L 15/24**
H04B 3/00, H04R 5/02, H04N 7/14, B60R 11/02, H05K 11/02, G10L 17/00, G10L 15/22, G10L 15/20

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden.

(71) Patentinhaber:
DaimlerChrysler AG, 70567 Stuttgart, DE

(72) Erfinder:
Fritzsche, Martin, Dipl.-Geophys., 88400 Biberach, DE; Kaltenmeier, Alfred, Dr., 89075 Ulm, DE; Linhard, Klaus, Dr.-Ing., 89601 Schelklingen, DE; Löhlein, Otto, Dipl.-Ing., 89171 Illerkirchberg, DE; Gloger, Joachim, Dipl.-Inform., 89346 Bibertal, DE; Schwarz, Tilo, Dr., 89075 Ulm, DE

(56) Für die Beurteilung der Patentfähigkeit in Betracht gezogene Druckschriften:
DE 199 62 218 A1
DE 199 48 546 A1
DE 199 42 868 A1
DE 691 01 527 T2

(54) Bezeichnung: **Bildgestützte adaptive Akustik**

(57) Hauptanspruch: Visuell-akustische Anordnung für Audiowiedergabe, Spracheingabe und Kommunikation zwischen mehreren Teilnehmern,
bei welcher Mikrofon- und/oder Lautsprecheranordnungen für die Teilnehmer individuell adaptiert sind,
und welche ein Kameraanordnung mit Bildverarbeitung einschließt, durch welche zumindest ein Teil des Bereichs, in welchem sich Teilnehmer befinden können erfasst wird,
dadurch gekennzeichnet,
dass ein Mittel vorgesehen ist, um die Mikrofon- und/oder Lautsprecheranordnungen auf Grundlage der Daten der Bildverarbeitung auf wenigstens einen der Teilnehmer auszurichten.

Beschreibung

[0001] Die Erfindung betrifft eine Anordnung und ein Verfahren zur bildgestützten adaptiven Akustik nach dem Oberbegriff der Patentansprüche 1 und 9.

[0002] Die Erfindung findet Verwendung bei Kommunikationssystemen insbesondere im Fahrzeug.

[0003] In einem Fahrzeugs sind die Sitzplätze fest installiert und dadurch ist die Position der möglichen Insassen relativ gut bestimmbar. Eine weitere Besonderheit ist, daß der Fahrer eine besondere Rolle unter den Insassen einnimmt. Jedoch ist es mit üblichen akustischen Mitteln nicht möglich, einen Hörer zu detektieren, wenn der Hörer passiv ist.

[0004] Gegebenenfalls wird eine Beschallung eines Sitzplatzes durchgeführt, obwohl kein Insasse vorhanden ist. Das führt zu unnötigen akustischen Störung

Stand der Technik

[0005] In der US Patentschrift 5,901,978 ist ein Verfahren und eine Vorrichtung zur Erfassung eines Kindersitzes angegeben, bei dem durch Mustererkennungsverfahren u.a. die Sitzbelegung erkannt wird und diese Information zur Steuerung z.B. der Sitzeinstellung, des Airbag Systems und des Unterhaltungssystems verwendet wird. Es werden visuelle und akustische Verfahren eingesetzt. Eine gezielte Geräuschreduzierung ist bei diesem System nicht vorgesehen.

[0006] Aus der Schrift DE 69101527 T2 ist eine Spracherkennungseinrichtung bekannt, welche in einem Fahrzeug verwendet wird. Die Vorrichtungen dient dazu, normalerweise von einem Fahrer durchgeführte manuelle Operationen zu ersetzen. Insbesondere bezieht sich diese Erfindung auf eine Spracherkennungsvorrichtung, die Befehle von einem Fahrersitz und einem Beifahrersitz lokalisiert und die Spracherkennungsrate in einem geräuscherfüllten Fahrzeugraum verstärkt, um dadurch die Zuverlässigkeit der Vorrichtung zu verbessern.

[0007] Die Schrift DE 19962218 A1 beschreibt ein System zum Autorisieren von Sprachbefehlen. Hierbei werden Sprachbefehle dadurch autorisiert, dass vorbestimmten Sprachbefehlen vorbestimmten Orte zugeordnet werden. An diesen Orten muss sich eine die Befehle sprechende Person befinden, damit der entsprechende Sprachbefehl ausgeführt wird. Der Sprachbefehl wird von einem Mikrofon erfasst. Gleichzeitig wird von einer dem vorbestimmten Ort zugeordneten Kamera die Mundbewegung einer dort sprechenden Person erfasst und der Befehl zur Ausführung nur freigegeben, wenn die Mundbewegung mit dem von Mikrofon erfassten Audiosignal korreliert.

[0008] Ein Verfahren und eine Vorrichtung zur Überprüfung Nutzungsberechtigung für ein Kommunikationssystem ist aus der Schrift DE 199 48 546 A1 bekannt. Hierbei werden die Gesichtszüge eines Kommunikationsteilnehmers erkannt und mit einem abgespeicherten Referenzbild verglichen. Mit dem Referenzbild ist desweiteren eine Sprachprobe abgespeichert, welche mit einer Kennungsphrase verglichen wird. Auf Grund dieser Vergleiche, wird ermittelt, ob ein aktueller Teilnehmer zur Nutzung des Systems berechtigt ist oder nicht.

Aufgabenstellung

[0009] Die Aufgabe der Erfindung besteht darin, eine Vorrichtung und ein Verfahren anzugeben, bei dem die Detektion der Sprachaktivität vereinfacht, die Adaption mit dem akustischen Signal durch Bild verbessert und bei starken und/oder instationären Geräuschen insbesondere im Fahrzeug Erkennungsfehler vermieden werden.

[0010] Die Erfindung betreffend die Anordnung ist in Anspruch 1 und betreffend das Verfahren in Anspruch 9 beschrieben. Vorteilhafte Ausgestaltungen und Weiterbildungen sind den Unteransprüchen zu entnehmen.

[0011] Mikrofon/Lautsprecheranordnungen gemäß der Erfindung sind geeignet, um mehrere Mikrofone oder Mikrofonarrays besser zu adaptieren, um bereits vor Sprachaktivität eine Initialisierung des Arrays auf den Sprecher zu erhalten. Wird die akustische Information und die Bildinformation fusioniert, so wird auch bei Sprachinaktivität die Erkennung und Identifizierung des Sprechers möglich. Dies ist für das Freisprechen, für die Spracherkennung und für die Insassen-Kommunikations-Systeme insbesondere im Fahrzeug vorteilhaft.

Ausführungsbeispiel

[0012] Die Erfindung wird im folgenden anhand von Ausführungsbeispielen beschrieben.

[0013] Durch die vordefinierten Sitzplätze werden neben der üblichen Verwendung eines akustischen Signals, z.B. eines einzelnen Mikrofons oder eines Mikrofon-Arrays zur Erfassung des Sprachsignale der Sprecher, verteilte Mikrofon-Arrays eingesetzt. Für jeden möglichen Insassen wird an dessen Sitzplatzposition in der Nähe seiner Mundposition ein Mikrofon plaziert. Es werden mehrere individuelle Insassen-Mikrofone eingesetzt. Jedes einzelne der Insassen-Mikrofone wird zur weiteren Steigerung der akustischen Qualität durch ein Mikrofonarray ersetzt. Es wird eine Anordnung von mehreren individuellen Insassen-Mikrofon-Arrays gebildet.

[0014] Dies hat den Vorteil, daß der Sprecher leicht zu detektieren ist, da durch die Nähe der individuellen Mikrofone garantiert ist, daß das lauteste Mikrofon-Signal, bzw. das lauteste Mikrofon-Array-Signal den aktiven Sprecher definiert.

[0015] Ein weiterer Vorteil bei der Verwendung individueller Insassen-Mikrofon-Arrays besteht darin, daß der Winkel des Sprechers zum Array bestimmt wird und dadurch die Sprecherposition genauer ermittelt wird.

[0016] Weiterhin ist vorteilhaft, daß durch die Verknüpfung der einzelnen Insassen-Mikrofone oder Arrays mit einer Sprecherverifikation die Identität der einzelnen Insassen und deren Sitzplatz erkannt wird.

[0017] Durch die Verküpfung der einzelnen Insassen-Mikrofone oder Arrays mit einem Spracherkenner sind vorteilhafterweise sprachbediente Operationen (Telefonbedienung, Radiobedienung u.ä.) von den Insassen ausführbar.

[0018] Durch die vordefinierten Sitzplätze wird neben der üblichen Verwendung eines einzelnen Lautsprechersystem für alle Insassen ein verteiltes Lautsprecher-Arrays eingesetzt, bei dem für jeden möglichen Insassen an dessen Sitzplatzposition in der Nähe seiner Ohrposition ein Lautsprecher platziert ist. Damit ergibt sich eine Anordnung von mehreren individuellen Insassen-Lautsprechern oder Arrays. Durch die Verwendung von Insassen-Lautsprecherarrays wird die akustische Qualität gesteigert. Es wird eine Anordnung von mehreren individuellen Insassen-Lautsprecher-Arrays gebildet.

[0019] Wird zusätzlich zur akustischen Erkennung die Erkennung durch die Bildverarbeitung mit eingeführt, tragen folgende Vorteile der Bilderkennung mit zur Insassenerkennung bei:

a) Die Bildverarbeitung erkennt, wieviele Insassen vorhanden sind, bzw. welche der Sitzplätze belegt sind.

b) Die Bildverarbeitung erkennt die Kopfposition der Insassen, die Ohren und den Mund.

c) Die Bildverarbeitung erkennt die Insassen (Insassen-Identifizierung).

[0020] Durch die Fusion von Sprache und Bild wird der Nachteil beseitigt, daß mit der Sprache eine Identifizierung des Insassen nur gelingt, wenn der Insasse spricht. Durch die zusätzliche Identifikation über Bild, ist die Identifizierung des Insassen immer möglich.

[0021] Insbesondere erfordert die Spracheingabe mit Mikrofon-Arrays die adaptierte Ausrichtung der Mikrofone auf den Sprecher, speziell auf den Mund des Sprechers. Mit dem akustischen Signal erfolgt die Ausrichtung, wenn der Insasse spricht.

[0022] Durch Kombination Sprache mit Bild ergeben sich folgende Vorteile:

– Initialisierung des Mikrofon-Arrays bevor der Insasse spricht. Dadurch ist bei einsetzender Sprache eine gute Start-Sprachqualität vorhanden. Auch die Detektion der Sprachaktivität wird vereinfacht, da der Sprach-Detektionsalgorithmus von einem Mikrofon-Array einfacher ist als der Bild-Detektionsalgorithmus.

– Bei Sprachaktivität wird die Adaption mit dem akustischen Signal durch Bild verbessert.

– Individuelle Mikrofone oder Mikrofon-Arrays für unbelegte Sitzplätze werden geschlossen. Dadurch werden fehlerhafte Sprachdetektionen abgeschaltet. Bei starken und/oder instationären Geräuschen im Fahrzeug wird durch das Abschalten eines Mikrofons ein deutlicher Vorteil erreicht. Es ergeben sich weniger Erkennungsfehler für den Fall, daß Geräusche aus dem nicht mit einem Insassen belegten Mikrofonsystem dem Spracherkenner angeboten werden.

[0023] Die Wiedergabe mit individuellen Lautsprechersystemen erfordert eine Ausrichtung auf die Ohren des Hörers. Mit dem akustischen Signal, dem Mikrofonsignal, erfolgt die Ausrichtung der Lautsprecher, wenn der Insasse spricht.

[0024] Durch Kombination Sprache mit Bild ergeben sich folgende Vorteile bei der Audio Ausgabe:

– wenn der Hörer hört und nicht spricht, erfolgt die Kopf/Ohr-Erkennung nur mit einer Bildverarbeitung.

– Nicht belegte Sitzplätze werden nicht beschallt. Dadurch entstehen keine unnötigen akustische Störungen der weiteren Insassen.

[0025] Mit den erfindungsgemäßen Anordnungen wird bei

1) individuellen Audio-Wiedergabesystemen die Information vorgegeben, welche Sitzplätze belegt sind. Nicht belegte Sitzplätze werden nicht beschallt. Bei belegten Sitzplätzen wird die Bildinformation benutzt um dem Kopf/Ohren zu folgen.

2) individuellen Sprach-Eingabesystemen die Information vorgegeben, welche Sitzplätze belegt sind, um die Mikrofone der nichtbelegten Sitzplätze abzuschalten.

[0026] Weitere Anwendungen findet die erfindungsgemäße Anordnung bei Insassen-Kommunikations-Systemen mit Spracherkennug und Sprechererkennung insbesondere im Fahrzeug. Das System aus Sprach- und Bildverarbeitung erkennt die Sitzbelegung, d.h. den Namen der Personen die sich auf den Sitzen befinden. Die Erkennung der Personen erfolgt durch Sprecherindentifizierung und/oder Gesichts-Identifizierung. Per Spracherkennung wird dann z.B. von einem Insassen gesagt: „Ich möchte mit Peter sprechen". Das System erkennt von wel-

chem Sitz gesprochen wird und erkennt auch den Sitzplatz von Peter. Es wird dann lediglich das Lautsprecher-Mikrofonsystem zwischen den beiden Personen aktiviert, die weiteren Personen werden nicht hinzugeschaltet und damit nicht gestört.

[0027] Sofern das Fahrzeug mit Monitoren an den einzelnen Sitzplätzen ausgestattet ist, kann das Gesicht des Sprechenden an den oder die Hörer gesendet werden.

[0028] Bei einer Videokonferenz mit Teilnehmern außerhalb des Fahrzeugs wird das Bild des jeweils Sprechenden mit übertragen. Falls jeder Sitzplatz mit einem Monitor ausgestattet ist, sehen die Teilnehmer innerhalb des Fahrzeugs jeweils den Sprechenden auf ihrem individuellen Monitor.

[0029] Die Erfindung ist nicht auf die angegebenen Ausführungsbeispiele beschränkt, sondern es ist die Verwendung in Konferenzsystemen jeglicher Art möglich.

Patentansprüche

1. Visuell-akustische Anordnung für Audiowiedergabe, Spracheingabe und Kommunikation zwischen mehreren Teilnehmern,
bei welcher Mikrofon- und/oder Lautsprecheranordnungen für die Teilnehmer individuell adaptiert sind,
und welche ein Kameraanordnung mit Bildverarbeitung einschließt, durch welche zumindest ein Teil des Bereichs, in welchem sich Teilnehmer befinden können erfasst wird,
dadurch gekennzeichnet,
dass ein Mittel vorgesehen ist, um die Mikrofon- und/oder Lautsprecheranordnungen auf Grundlage der Daten der Bildverarbeitung auf wenigstens einen der Teilnehmer auszurichten.

2. Visuell-akustische Anordnung nach Anspruch 1, dadurch gekennzeichnet, daß die Anordnung in einem Fahrzeug mit vordefinierten Sitzplätze eingebaut ist.

3. Visuell-akustische Anordnung nach Anspruch 1 und 2, dadurch gekennzeichnet, daß zur Erfassung der Sprachsignale der Teilnehmer, verteilte Mikrofon-Arrays derart angebracht sind, daß für jeden möglichen Teilnehmer an dessen Position in der Nähe seiner Mundposition zumindest ein Mikrofon oder Mikrofonarray plaziert ist.

4. Visuell-akustische Anordnung nach Anspruch 1 und 2, dadurch gekennzeichnet, daß für alle Teilnehmer ein verteiltes Lautsprecher-Arrays eingebaut ist, bei dem für jeden möglichen Teilnehmer an dessen Position in der Nähe seiner Ohrposition ein Lautsprecher oder Lautsprecherarray platziert ist.

5. Visuell-akustische Anordnung nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Mikrofonanordnung mit einem Spracherkennungs- und/oder Sprecheridentikationssystem verbunden ist.

6. Visuell-akustische Anordnung nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Mikrofon/Lautersprecher-Arrays jedes einzelnen Teilnehmers individuell ein- und ausschaltbar sind.

7. Visuell-akustische Anordnung nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß zusätzlich zur akustischen Erkennung der Teilnehmer eine Erkennung über die Bildverarbeitung erfolgt zur Bestimmung der Kopfposition, der Ohren und des Mundes und zur Identifizierung der Teilnehmer.

8. Visuell-akustische Anordnung nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß ein Monitor für jeden Teilnehmer zuschaltbar ist, auf dem die Teilnehmer sichtbar sind.

9. Visuell-akustisches Verfahren für Audiowiedergabe, Spracheingabe und Kommunikation zwischen mehreren Teilnehmern,
dass Mikrofon- und/oder Lautsprecheranordnungen für die Teilnehmer individuell adaptierbar sind,
und dass eine Kameraanordnung mit Bildverarbeitung, zumindest einen Teil des Bereichs, in welchem sich Teilnehmer befinden können erfasst,
dadurch gekennzeichnet,
dass auf der Grundlage der Erkennung der Teilnehmer mittels Mikrofon- und/oder Lautsprecheranordnungen auf wenigstens einen der Teilnehmer ausgerichtet werden.

10. Verfahren nach Anspruch 9, dadurch gekennzeichnet, daß durch die verteilte, individuelle Anordnung der aktive Sprecher durch das lauteste Mikrofon-Signal oder das lauteste Mikrofon-Array-Signal definiert wird.

11. Verfahren nach Anspruch 9, dadurch gekennzeichnet, daß bei der Verwendung individueller Insassen-Mikrofon-Arrays der Winkel des Sprechers zum Array bestimmt wird und dadurch die Sprecherposition genauer ermittelt wird.

12. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß durch die Verknüpfung der einzelnen Insassen-Mikrofone-(Arrays) mit einer Sprecherverifikation die Identität der einzelnen Teilnehmer und deren Position erkannt wird.

13. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß durch die

Verküpfung der einzelnen Insassen-Mikrofone oder Arrays mit einem Spracherkenner sprachbediente Operationen von den Teilnehmern durchgeführt werden.

14. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß durch die Kombination der akustischen Teilnehmererkennung mit der Bildverarbeitung eine Initialisierung des Mikrofon-Arrays durchgeführt wird, bevor der Teilnehmer spricht.

15. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß bei Sprachaktivität die Adaption mit dem akustischen Signal durch das Bild verbessert wird.

16. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß mit dem akustischen Signal die Ausrichtung der Lautsprecher oder Lautsprecherarrays erfolgt, wenn der Teilnehmer spricht.

17. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß wenn der Hörer hört und nicht spricht, die Kopf/Ohr-Erkennung nur mit einer Bildverarbeitung durchgeführt wird.

18. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß mittels Bildverarbeitung nicht belegte Sitzplätze erkannt werden, und daß die individuellen Mikrofone oder Mikrofon-Arrays für unbelegte Sitzplätze geschlossen werden und dadurch fehlerhafte Sprachdetektionen abgeschaltet werden.

19. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass bereits vor einer Sprachaktivität zumindest eines Teilnehmers eine Initialisierung des Arrays auf den Teilnehmer erfolgt.

Es folgt kein Blatt Zeichnungen