

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 1 251 493 A2

(12)

EUROPÄISCHE PATENTANMELDUNG

(43) Veröffentlichungstag:

23.10.2002 Patentblatt 2002/43(51) Int Cl.7: **G10L 21/02**(21) Anmeldenummer: **02008011.5**(22) Anmeldetag: **10.04.2002**

(84) Benannte Vertragsstaaten:

**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL PT SE TR**

Benannte Erstreckungsstaaten:

AL LT LV MK RO SI(30) Priorität: **14.04.2001 DE 10118653**

(71) Anmelder:

- **DaimlerChrysler AG**
70567 Stuttgart (DE)
- **TEMIC Sprachverarbeitung GmbH**
89077 Ulm (DE)

(72) Erfinder:

- **Buck, Markus**
88499 Zwiefaltendorf (DE)

• **Haulick, Tim, Dr.****89134 Blaustein (DE)**• **Linhard, Klaus, Dr.****89601 Schelkingen (DE)**(74) Vertreter: **Straub, Bernd**

DaimlerChrysler AG,
Intellectual Property Management,
Sedanstr. 10 / Geb. 17
89077 Ulm (DE)

Bemerkungen:

Ein Antrag gemäss Regel 88 EPÜ auf Berichtigung (Vervollständigung des Anspruchs 1) liegt vor. Über diesen Antrag wird im Laufe des Verfahrens vor der Prüfungsabteilung eine Entscheidung getroffen werden (Richtlinien für die Prüfung im EPA, A-V, 3.).

(54) **Verfahren zur Geräuschreduktion mit selbststeuernder Störfrequenz**

(57) Die Erfindung betrifft ein Verfahren, mit dem Sprache in gestörter Umgebung mit möglichst hoher Sprachqualität erfaßt wird. Dazu wird ein kompaktes Array aus beispielsweise zwei Einzelmikrofonen durch Signalverarbeitungsverfahren aus adaptivem Strahlfor-

men und spektraler Subtraktion zu einem System zusammengefaßt. Das Referenzsignal des Strahlformers wird durch Kombination mit einer spektralen Subtraktion von Sprachsignalanteilen soweit befreit, daß ein Referenzsignal der Störung entsteht und der Strahlformer hohe Gewinne ergibt.

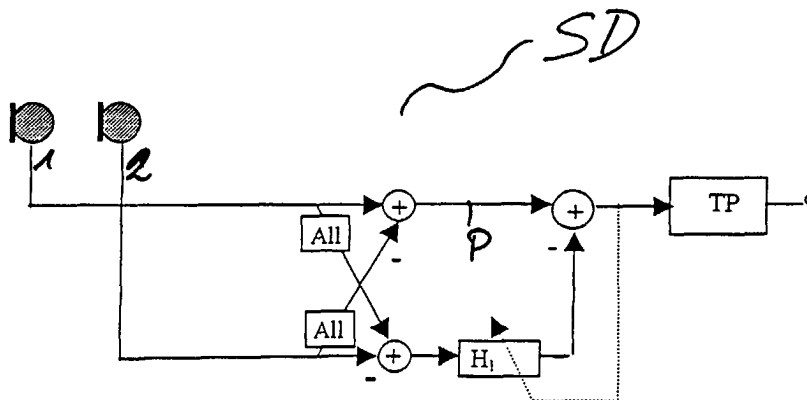


Fig.1

EP 1 251 493 A2

Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zur Geräuschreduktion nach dem Oberbegriff des Patentanspruchs 1.

Ein häufig verwendetes Verfahren zur Geräuschreduktion eines gestörten Nutzsignals, z.B. ein Sprachsignal, Musiksinal etc. ist die spektrale Subtraktion. Vorteil der spektralen Subtraktion ist die geringe Komplexität und daß das gestörte Nutzsignal nur in einer Variante (nur 1 Kanal) benötigt wird. Nachteil ist die Signalverzögerung (bedingt durch die Blockverarbeitung im Spektralbereich), die begrenzte maximal erreichbare Geräuschreduktion und die Schwierigkeit instationäre Geräusche zu kompensieren. Stationäre Geräusche können bei noch guter Sprachqualität z.B. um 12dB reduziert werden.

[0002] Wird eine höhere Geräuschreduktion oder eine bessere Sprachqualität verlangt, sind mehrere Aufnahmekanäle erforderlich. Es werden z.B. Mikrofon-Arrays verwendet. Von den verschiedenen Mikrofon-Arrays sind für viele praktische Anwendungen solche besonders interessant, die mit kleinen geometrischen Abmessungen für die Mikrofonanordnung auskommen. Es werden kleine differentielle Mikrofon-Arrays (auch superdirektive Arrays genannt) gebildet und eine adaptive Variante dieser Mikrofonanordnung, wobei zur Adaption der LMS (least mean square) -Algorithmus verwendet wird. Bei der adaptiven Form dieses Arrays werden zwei Mikrofone laufzeitkompensiert auf zwei Arten derart subtrahiert, daß ein virtuelles Mikrofon mit nierenförmiger Richtcharakteristik zum Sprecher und ein 'virtuelles' Mikrofon mit nierenförmiger Charakteristik vom Sprecher abgewandt entsteht. Die Laufzeitkompensation entspricht der Zeit, die der Schall für die Distanz zwischen den beiden Mikrofonen benötigt, z.B. 1,5cm. Es ergibt sich eine "Rücken-an-Rücken" nierenförmige Richtcharakteristik. Das zum Sprecher gerichtete Mikrofon ist das primäre Signal für das adaptive Filter und das entgegengesetzt gerichtete Mikrofon ist das Referenzsignal der Störung.

[0003] Figure 1 zeigt eine adaptive Anordnung für einen Strahlformer. Der Laufzeitausgleich mit einem Allpaß ALL wird durch Verschiebung um ganze Abtastwerte realisiert. Durch die oben beschriebene Kombination zweier Einzelmikrofone mit Kugelcharakteristik ergibt sich eine nierenförmige Richtcharakteristik zum Sprecher und eine entgegengesetzt gerichtete nierenförmige Richtcharakteristik als Störreferenz. Das adaptive Filter H 1 wird im Zeitbereich mit dem LMS (least mean square) -Algorithmus adaptiert. Ein Tiefpaß TP am Systemausgang hebt tiefe Frequenzanteile an, die bei der Bildung der nierenförmigen Richtcharakteristik gedämpft werden.

Die Anordnung der Mikrofone M hintereinander gemäß Figur 1 wird als 'end fire array' bezeichnet, im Gegensatz wird die Anordnung der Mikrofone nebeneinander mit 'broad side array' bezeichnet.

[0004] Figure 2 zeigt eine Anordnung für ein "broad side array" aus zwei Mikrofonen im Abstand, wobei mit Hilfe der spektralen Subtraktion (SPS) die beiden Mikrofonensignale vorverarbeitet werden. Ein Laufzeitausgleich mit dem Allpaß All zwischen beiden Kanälen wird ausgeführt und dient dem Ausgleich von Bewegungen des Sprechers. Die Summe der beiden vorverarbeiteten Mikrofonensignale bildet den primären Eingang und die Differenz den Referenzeingang für ein adaptives Filter H 1. Das adaptive Filter in dieser Anordnung mit Summen- und Differenzeingang wird auch als 'generalized sidelobe canceller' bezeichnet. Die Adaption erfolgt mit dem LMS-Algorithmus, wobei die Implementierung des LMS im Frequenzbereich erfolgt. Eine Nachverarbeitung der Mikrofonensignale wird mit einer modifizierten Kreuzkorrelationsfunktion im Frequenzbereich durchgeführt. Die grundlegende Struktur mit spektraler Vorverarbeitung mittels SPS, Strahlformung und Nachverarbeitung (Post) ist in der Patentschrift EP 0615226B1 beschrieben, wobei eine genaue Spezifizierung des Strahlformers nicht erfolgt ist.

[0005] Figure 3 zeigt einen Überblick über Schaltungsanordnungen von Mikrofonen zur Bildung der Richtcharakteristiken für zwei Mikrofone. Die beiden einzelnen Mikrofone selbst können bereits eine nierenförmige Charakteristik haben oder die sogenannte Kugelcharakteristik. "ALL" bezeichnet einen Allpaß für den Laufzeitausgleich. 'Gain' ist ein Verstärkungsausgleich zwischen beiden Kanälen der in der Praxis erforderlich ist, um die Empfindlichkeit der Mikrofonkapseln anzugleichen.

[0006] Die Einsprechrichtung in den Polardiagrammen der Richtcharakteristiken ist 90°. Die ersten 3 Anordnungen a, b und c sind als Sprachkanal geeignet, da bei 90° ein Maximum vorliegt und für die weiteren Richtungen eine Dämpfung vorhanden ist. Anordnung a und b führen auf die gleiche Richtcharakteristik. Die Anordnungen a, b werden als Summen- oder Differenz Array und Anordnung c als differentielles Array bezeichnet. Die Anordnungen d und e haben eine Nullstelle bei 90° im Polardiagramm und sind damit als Störreferenz geeignet. Die Nullstelle bei 90° im Polardiagramm ist notwendig, damit keine Sprachanteile in den Referenzkanal gelangen. Sprachanteile im Referenzkanal führen zur teilweisen Kompensation der Sprache.

Unter idealen Bedingungen wird sich gemäß Anordnung d und e für die Störreferenz eine Nullstelle in Richtung zum Sprecher einstellen. In praktischen Anwendungen wird dies jedoch nicht der Fall sein. Die Folge ist, daß Sprachanteile wie Störsignale behandelt werden und damit vom eigentlichen Sprachsignal entfernt werden.

[0007] Strahlformer werden meist nur in den Sprachpausen adaptiert, um keine Adaption an Sprachanteile zu ermöglichen. Dennoch werden auch in diesem Fall in der Referenz vorhandene Sprachanteile kompensiert, da sie dem Geräusch stets überlagert sind.

[0008] Eine andere Vorgehensweise ist die Verstär-

kung von Kanälen anzugleichen, damit bei deren Subtraktion im Idealfall eine Nullstelle erzeugt wird. Dies ist notwendig, da Mikrofone aus der Serienfertigung Toleranzen ausweisen. In den Anordnungen der Figur 3 ist dies mit dem Funktionsblock 'Gain' berücksichtigt, der unterschiedliche Mikrofon-Empfindlichkeiten ausgleicht.

[0009] In Anwendungen wird trotz Empfindlichkeitsausgleich mit 'Gain' dennoch keine Nullstelle für das Sprachsignal in der Referenz eingestellt. Nur unter der Voraussetzung, daß das Mikrofon im akustischen Freifeld betrieben wird (ohne Reflexionen), können die Sprachanteile vollständig kompensiert werden. Reale Anwendungen haben bedingt durch Reflexionen einen gewissen Schallanteil aus unterschiedlichen Richtungen, der eine Nullstelle für das Sprachsignal nicht entstehen läßt. Es wird sich bei Anordnungen gemäß Figur 1 oder Figur 2 stets ein gewisser Sprachanteil in dem Referenzsignal des Strahlformers wiederfinden, der zu Sprachverzerrungen führt.

[0010] Der vorliegende Erfindung liegt deshalb die Aufgabe zugrunde, ein Verfahren zur Geräuschreduktion anzugeben, mit dem ein Übersprechen des Nutzsignals in das Störreferenzsignal minimiert wird.

[0011] Die Erfindung ist in Anspruch 1 angegeben. Vorteilhafte Ausgestaltungen und Weiterbildungen sind den Unteransprüchen zu entnehmen.

[0012] Die Erfindung hat den Vorteil, daß deutlich weniger Nutzsignalanteile, z.B. Sprachanteile im Störreferenzsignal vorhanden sind als mit den bisherigen Verfahren. Die Beseitigung der störenden Sprachanteile ist damit unter realen Bedingungen mit Reflexionen des Sprachsignals in realen Räumen wie z.B. im Kraftfahrzeug möglich.

[0013] Die Erfindung geht davon aus, daß zur Bildung des Störreferenzsignals eine einseitige spektrale Subtraktion durchgeführt wird. Wesentlich ist, daß die spektrale Subtraktion zur Bildung eines Referenzsignals nur an einem Kanal stattfindet, was mit 'einseitig' bezeichnet wird. Der eine Kanal enthält damit Nutz- und Störsignale, der zweite Kanal nach der spektralen Subtraktion enthält nur Nutzsignale. Bei der anschließenden Subtraktion der beiden Kanäle wird der Nutzanteil subtrahiert und es verbleibt die Störung. Diese Differenz ist das Störreferenzsignal.

[0014] Werden z.B. Mikrofone zur Aufnahme von Sprachsignalen verwendet, so werden die Sprachsignale derart verarbeitet, daß das Störreferenzsignal eine Nullstelle zum Sprecher in der Form einer nierenförmigen oder einer achtförmigen Charakteristik aufweist. Die einseitige spektrale Subtraktion führt zu einer selbststeuernden Regelung der Charakteristik, derart, daß die Nullstelle nur bei Sprachaktivität entsteht. In Sprachpausen führt die einseitige spektrale Subtraktion dazu, daß nichts oder nur ein geringes Signal subtrahiert wird und damit näherungsweise die Charakteristik des Einzelmikrofons (z.B. Niere oder Kugel) für die Störung zur Verfügung steht.

[0015] Die ideale Nullstelle für das Sprachsignal in der Referenz wird nur mit einer idealen spektralen Subtraktion im akustischen Freifeld erreicht. Eine ideale spektrale Subtraktion ergibt das ungestörte Sprachsignal als Ausgangssignal und würde dann jede weitere Bearbeitung unnötig machen. Die spektrale Subtraktion in der Praxis ergibt nur eine gute Annäherung des Sprachsignals mit Geräuschresten in den Sprachpausen. Da die einseitige spektrale Subtraktion ergänzend zu der Mikrofon-Nullstelle eingesetzt wird, vermindern sich die Sprachanteile der Referenz deutlich.

[0016] Das Restgeräusch der spektralen Subtraktion in Sprachpausen wird mit einem Parameter eingestellt, dem 'spectral floor'. Der spectral floor b ist der minimale Wert eines Filterkoeffizienten W der spektralen Subtraktion bei jedem Frequenzindex i . Das Ausgangssignal $Y(i)$ ergibt sich durch Multiplikation der Filterkoeffizienten $W(i)$ mit dem Eingangswert $X(i)$:

$$W(i) = \max (W(i), b);$$

und

$$Y(i) = W(i) \cdot X(i);$$

[0017] Der maximale Wert für W ist 1 (Ausgang = Eingang). Wird $b=1$ gewählt, ist die spektrale Subtraktion praktisch ausgeschaltet. Mit $b=0$ erreicht die spektrale Subtraktion die maximale Wirksamkeit. In der Praxis ergibt sich mit $b=0$ eine schlechte Sprachqualität.

[0018] Mit dem Parameter b ergibt sich für die vorliegende Erfindung die Möglichkeit die einseitige spektrale Subtraktion in ihrer Wirksamkeit kontinuierlich einzustellen. Mit einem Wert von z.B. $b=0.25$ wird eine Geräuschunterdrückung von ca. 12dB und eine gute Sprachqualität erzielt.

[0019] Die Erfindung wird im folgenden anhand von Ausführungsbeispielen unter Bezugnahme auf schematische Zeichnungen näher erläutert.

[0020] Figur 4 zeigt 3 Blockschaltbilder mit einseitiger spektraler Subtraktion für den Referenzeingang. In Figur 4a ist das primäre Nutzsignal P des Strahlformers (z. B. Sprachsignal) als differentiell Array DA für die Kanäle 1, 2 geschaltet ist (Anordnung c in Figur 3). Figur 4b, 4c zeigt eine Schaltung des Primärsignals P als Summen- und Differenz Array SD (Anordnung a und b in Figur 3).

Der Störreferenzeingang verarbeitet das Referenzsignal R mit der zusätzlichen Erweiterung der einseitigen spektralen Subtraktion in differentieller Form gemäß den Anordnung d und e in Figur 3. Die Differenz aus Nutzsignal in Kanal 2 und entstörtem Nutzsignal aus Kanal 1 wird auf das adaptive Filter H_1 gegeben. Das adaptive Filter H_1 wird im Zeitbereich oder in einer äquivalenten Form im Frequenzbereich mit dem LMS - Algorithmus adaptiert. Das gefilterte Störreferenzsignal R

wird anschließend vom primären Nutzsignal P subtrahiert.

[0021] Eine weitere Ausgestaltung der Erfindung gemäß Figur 5 besteht darin, daß die einseitige spektrale Subtraktion, SPS₁, einmal am Kanal 1 für das Nutzsignal durchgeführt wird, um damit zusammen mit dem Nutzsignal in Kanal 2 ein erstes Referenzsignal R1 zu bilden. Ein zweites Mal wird die einseitige spektrale Subtraktion, SPS₂, am Nutzsignal des Kanal 2 durchgeführt, um zusammen mit dem Nutzsignal in Kanal 1 ein zweites Referenzsignal R2 zu bilden. Es entsteht ein System mit 2 Referenzsignalen, die vom Primärsignal P subtrahiert werden. Bei Sprachsignalen wird in den Sprachpausen die Störung jeweils mit der Charakteristik der Einzelmikrofone erfaßt und bei Sprachaktivität eine Nullstelle für das Sprachsignal erzeugt.

[0022] Entsprechend den Erläuterungen zu den Blockschaltbildern der Figur 4 wird die Abwandlung mit 2 Referenzeingängen für 'end fire' Mikrofonanordnung oder 'broad side' Anordnung verwendet. Figure 5 zeigt das Blockschaltbild für die 'end fire' Anordnung. Der Strahlformer besteht aus dem Kanal 1 für das Sprachsignal und zwei Referenzkanälen 2, 3. Jeder Referenzeingang wird von einem adaptiven Filter 'H₁', bzw. 'H₂' gefiltert. Der Filterabgleich erfolgt mit einem mehrkanaligen LMS-Algorithmus.

[0023] Stehen mehr als 2 Eingangssignale zur Verfügung, so wird durch Kombination von jeweils 2 Eingängen in der beschriebenen Weise eine einseitige spektrale Subtraktion durchgeführt, um ein Referenzsignal zu erhalten. Wird z.B. ein 'broad side array' mit 3 Mikrofonen angenommen, ergeben sich für die Paarbildung 6 Kombinationen. Wird berücksichtigt, daß bei jedem Paar die einseitige spektrale Subtraktion wahlweise bei dem einen oder dem anderen Kanal durchgeführt wird, so verdoppelt sich die Anzahl der Kombinationen und somit die Anzahl der Referenzkanäle. Bei einem Array aus mehreren Mikrofonen wird eine eingeschränkte Anzahl aus den möglichen Kombinationen verwendet.

Die Erfindung ist nicht auf die Aufzeichnung der Nutzsignale durch Mikrofone beschränkt, sondern es können Empfangssysteme wie z.B. Antennen verwendet werden. Nutzsignale können jegliche Art von akustischen und elektrischen Signalen sein.

Patentansprüche

1. Verfahren zur Erzeugung eines Störreferenzsignals R zur Geräuschreduktion eines primären Nutzsignals, welches durch Kombination der Signale wenigstens zweier Kanäle, insbesondere Sprachkanäle, gebildet wird, wobei die Signale paarweise miteinander verarbeitet werden, und wobei jeweils nur eines der paarweise verarbeiteten Signale einer spektralen Subtraktion unterzogen und zur Differenzbildung mit dem anderen Si-

gnal herangezogen wird,

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, daß** das primäre Nutzsignal als differentielles Array (DA) von zwei Kanälen (1, 2) geschaltet wird.
3. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, daß** das primäre Nutzsignal als Summen- und Differenz Array (SD) von zwei Kanälen (1, 2) geschaltet wird.
4. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, daß** das Störreferenzsignal mit der zusätzlichen Erweiterung der einseitigen spektralen Subtraktion in differentieller Form derart erzeugt wird, daß die Differenz aus dem entstörten Nutzsignal aus einem Kanal (1) und dem Nutzsignal aus einem weiteren Kanal (2) auf ein adaptives Filter (H1) gegeben wird, und daß das gefilterte Störreferenzsignal (R) anschließend vom primären Nutzsignal (P) subtrahiert wird.
5. Verfahren nach einem der Ansprüche 1 bis 3, **dadurch gekennzeichnet, daß** eine spektrale Subtraktion (SPS₁) an einem ersten Kanal (1) für das Nutzsignal durchgeführt wird und zusammen mit dem Nutzsignal in einem zweiten Kanal (2) auf ein adaptives Filter (H1) gegeben wird und ein erstes Referenzsignal (R1) gebildet wird, daß eine weitere spektrale Subtraktion (SPS₂) am Nutzsignal des zweiten Kanals (2) durchgeführt wird und zusammen mit dem Nutzsignal aus dem ersten Kanal (1) auf ein adaptives Filter (H2) in einem weiteren Kanal (3) gegeben wird und ein zweites Referenzsignal (R2) gebildet wird, und daß die beiden Referenzsignale (R1, R2) vom primären Nutzsignalsignal (P) subtrahiert werden.
6. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, daß** die Filter (H1, H2) im Zeitbereich oder im Frequenzbereich mit dem LMS - Algorithmus adaptiert werden.
7. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, daß** das Nutzsignal von Mikrofonen aufgezeichnet wird.
8. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, daß** als Nutzsignal ein Sprachsignal verwendet wird.
9. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, daß** die spektrale Subtraktion mit einem Parameter in ihrer Wirksamkeit kontinuierlich eingestellt wird.
10. Verfahren nach Anspruch 9, **dadurch gekenn-**

zeichnet, daß der Parameter als minimaler Wert eines Filterkoeffizienten der spektralen Subtraktion bei jedem Frequenzindex gebildet wird.

11. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, daß** bei mehr als zwei Eingangssignalen durch Kombination von jeweils zwei Eingängen eine spektrale Subtraktion zur Erzeugung eines Referenzsignals durchgeführt wird.

10

15

20

25

30

35

40

45

50

55

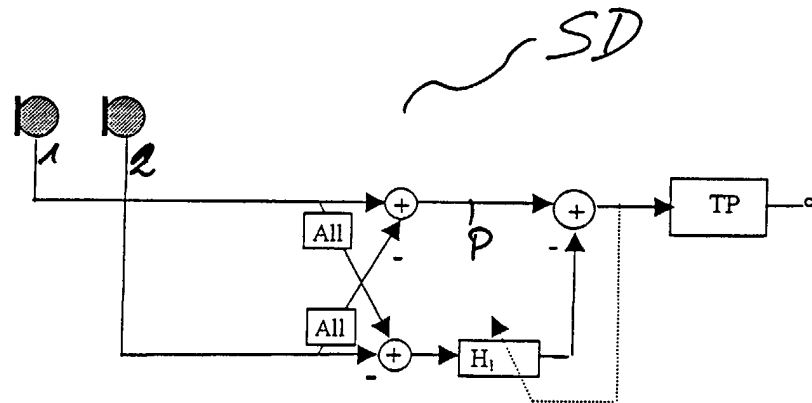


Fig.1

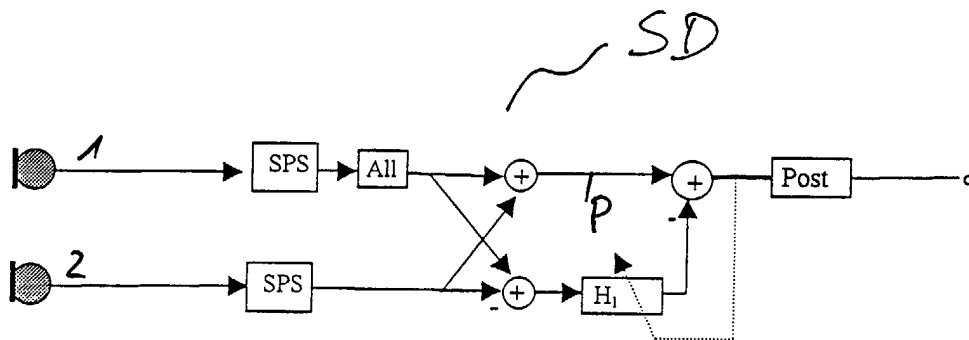
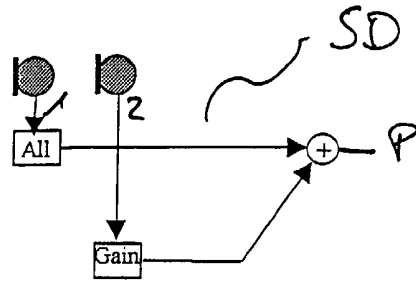
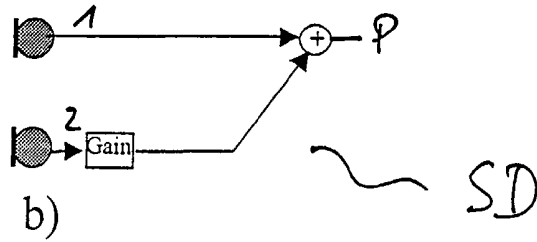


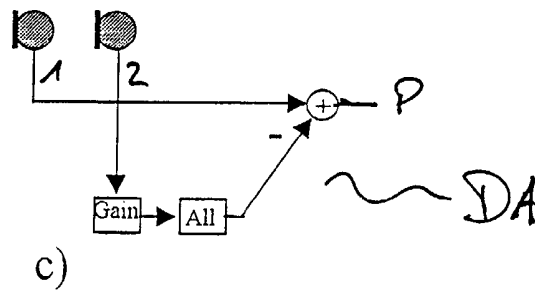
Fig.2



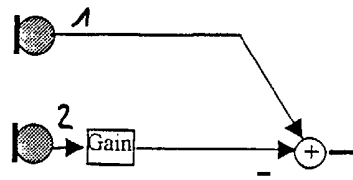
a)



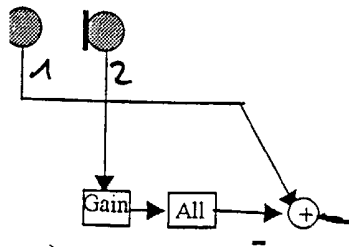
b)



c)



d)



e)

Fig. 3

