(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Veröffentlichungsnummer: **0 612 059 A2**

(12) # EUROPÄISCHE PATENTANMELDUNG

(21) Anmeldenummer: **93120010.9**

(51) Int. Cl.[5]: **G10L 3/00**, G10L 3/02

(22) Anmeldetag: **11.12.93**

(30) Priorität: **23.12.92 DE 4243831**

(43) Veröffentlichungstag der Anmeldung:
24.08.94 Patentblatt 94/34

(84) Benannte Vertragsstaaten:
DE FR GB

(71) Anmelder: **Daimler-Benz Aktiengesellschaft**
Postfach 80 02 30
D-70546 Stuttgart (DE)

(72) Erfinder: **Linhard, Klaus, Dr.-**
Kasernstrasse 43
D-89231 Neu-Ulm (DE)

(74) Vertreter: **Amersbach, Werner, Dipl.-Ing.**
AEG Aktiengesellschaft
Postfach 70 02 20
D-60591 Frankturt (DE)

(54) **Verfahren zur Laufzeitschätzung an gestörten Sprachkanälen.**

(57) Die Erfindung betrifft ein Verfahren zur Geräuschreduktion in einem Spracherkennungssystem. Es werden die Phasen von zumindest zwei gestörten Signalen geschätzt. Die Phasenschätzung und der für die Geräuschreduktion erforderliche Phasenausgleich wird im Frequenzbereich durchgeführt. Die Hintergrundstörung und das Einschwingverhalten des Raumes werden ständig mitgeschätzt.

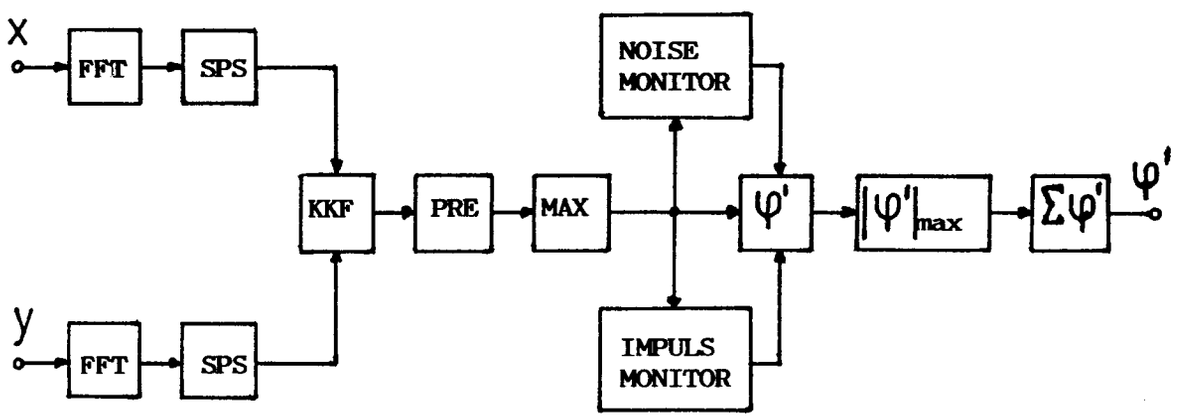

FIG.1

EP 0 612 059 A2

Die Erfindung betrifft ein Verfahren nach dem Oberbegriff des Patentanspruchs 1.

Ein derartiges Verfahren findet Verwendung bei automatischen Spracherkennungssystemen oder für Freisprechanlagen z.B. in Büroräumen, Kraftfahrzeugen etc..

Gestörte Sprache ist besser erfaßbar, wenn sie mit zwei oder mehreren Kanälen aufgezeichnet wird. Der Mensch benutzt zwei Kanäle, seine beiden Ohren. Durch eine psychoakustische Nachverarbeitung wird bei ihm die Richtung des Sprechers ermittelt und die Hintergrundstörung ausgeblendet. Bei technischen Geräten können zwei oder mehrere Kanäle zur Aufzeichnung verwendet werden. Diese Signale können dann mit einer digitalen Signalverarbeitung aufbereitet werden.

Ein wesentlicher Aspekt der mehrkanaligen Verarbeitung ist die Schätzung des Laufzeitunterschiedes der einzelnen Kanäle. Ist der Laufzeitunterschied bekannnt, kann die Richtung des Schallereignisses (Sprecher) ermittelt werden. Die Signale der einzelnen Kanäle können entsprechend laufzeitkorrigiert und weiterverarbeitet werden. Werden z.B. nicht korrigierte Signale zu einem Summensignal zusammengefaßt, können sich einzelne spektrale Anteile des Signals durch Interferenz verstärken, dämpfen oder auslöschen.

Ein Verfahren zur automatischen Ermittlung der Laufzeitunterschiede zweier Mikrofone ist aus einer Veröffentlichung von M. Schlang, ITG-Fachtagung 1988, Bad Nauheim S. 69-73 bekannt. Es arbeitet im Zeitbereich. Jedoch ist dieses Verfahren bei starken Störungen nicht anwendbar.

Der Erfindung liegt deshalb die Aufgabe zugrunde ein Verfahren zur Laufzeitschätzung für ein Spracherkennungssystem anzugeben, das auch bei starken Hintergrundgeräuschen anwendbar ist, für ein Mehrkanalübertragungssystem geeignet ist und zeit- und kostensparend arbeitet.

Die Aufgabe wird gelöst durch die im kennzeichnenden Teil des Patentanspruchs 1 angegebenen Merkmale. Vorteilhafte Ausgestaltungen und/oder Weiterbildungen sind den Unteransprüchen zu entnehmen.

Die Erfindung wird anhand eines Ausführungsbeispiels beschrieben unter Bezugnahme auf schematische Zeichnungen.

In FIG. 1 wird anhand eines Blockschaltbilds die Phasenschätzung erläutert.

FIG. 2 gibt für ein Fahrgeräusch von 140km/h eine Darstellung der Größen S_B, S_I, S_N und g in Abhängigkeit von der Zeit an.

In der vorliegenden Erfindung wird ein 2-kanaliger Laufzeitausgleich vorgestellt. Die Erweiterung auf mehrere Kanäle ist mit dem entsprechenden Mehraufwand leicht möglich. Der Laufzeitausgleich ist ein Teil der Signalvorverarbeitung einer mehrkanaligen Geräuschreduktion, die z.B. für einen Spracherkenner im Fahrzeug verwendet werden kann.

Die Laufzeit wird im Frequenzbereich ermittelt. Dies ermöglicht eine einfache Laufzeitkorrektur durch die Multiplikation des Spektrums mit der neuen Phase und führt zu einem geringen Rechenaufwand.

Die Sprach- und Geräuschaufnahmen zur Entwicklung und Bewertung des vorliegenden Verfahrens wurden in einem Fahrzeug mit zwei Mikrofonen durchgeführt. Die Störung ist das Fahrgeräusch bei verschiedenen Fahrsituationen.

Mit dem erfindungsgemäßen Verfahren werden im Frequenzbereich die Phasen an einer Anzahl von Maxima der Kreuzkorrelation bestimmt. Die Hintergrundstörung und das Einschwingverhalten des Raumes werden ständig mitgeschätzt. Die einzelnen Phasenwerte werden nur zu Beginn eines Einschwingvorgangs verarbeitet und wenn das Hintergrundgeräusch um einen gewissen Faktor überschritten wird. Bei der Weiterverarbeitung der Phasenwerte wird eine lineare Phasenbeziehung vorausgesetzt und die Varianz der Schätzung wird bei der Glättung der Werte mitberücksichtigt. Die Berücksichtigung des Einschwingvorgangs des Raumes fuhrt dazu, daß nur bei starken Energieanstiegen der Sprache eine Phasenschätzung stattfindet. Sofort zu Beginn des Wortes steht ein neuer Phasenschätzwert zur Verfügung. Der Einfluß von Reflexionen wird vermindert. Durch die Berücksichtigung des Hintergrundgeräuschs ist das Verfahren für den praktischen Einsatz z.B. im Fahrzeug gut geeignet. Anhand eines Blockschaltbildes in FIG. 1 wird der Verfahrensablauf der Phasenschätzung näher erläutert.

Die Mikrofonsignale x und y werden in den Frequenzbereich transformiert (FFT, Fast Fourier Transformation). Die Transformationslänge wird zu N = 256 gewählt. Es ergaben sich die transformierten Segment X_l (i) und Y_l (i). l bezeichnet den Blockindex der Segmente, i die diskrete Frequenz (i = 0,1,2,...,N-1). Die Segmente sind halb überlappt und werden mit einem Hanning Fenster gewichtet. (Die Abtastrate der Signale x und y beträgt 12 kHz.)

Im Frequenzbereich wird der Langzeitmittelwert des Betragsspektrums subtrahiert (SPS, spektrale Substraktion). Die Phase der Signale wird nicht verändert. Das Störgeräusch wird reduziert. Es ergeben sich die Schätzwerte $\hat{X}$ und $\hat{Y}$. Die SPS ist ein Standardverfahren und kann hier in einer einfachen Version eingesetzt werden. Sind nur geringe Störungen vorhanden, kann auf die SPS ganz verzichtet werden.

Mit der Glättungskonstante β wird das Störspektrum $S_{nn}(i)$ geschätzt. Das Störspektrum wird normiert und subtrahiert. l bezeichnet den Blockindex, i die diskrete Frequenz. Als Glättungskonstante wird z.B. β_l =

0.03 verwendet.

$$\hat{S}_{nn,l}(i) = (1-\beta_l)\hat{S}_{nn,l-1}(i) + \beta_l|X_l(i)|^2 \quad (1)$$

$$|\hat{X}_l(i)| = |X_l(i)| - \frac{\hat{S}_{nn,l}(i)}{|X_l(i)|} \quad (2)$$

$$\hat{X}_l(i) = \left[1 - \frac{\hat{S}_{nn,l}(i)}{|X_{l(i)}|^2}\right] X_l(i) \quad (3)$$

Für den zweiten Kanal Y gelten die entsprechenden Gleichungen.

Aus den geschätzten Werten $\hat{X}$ und $\hat{X}$ wird der Betrag der Kreuzleistungsdichte $B_{XY,l}$ berechnet. Der Bereich (N_u, N_o) liegt z.B. zwischen 300 und 1500Hz (N_u = 6, N_o = 31, bei N = 256). Dabei gilt

$$S_{xy,l}(i) = (1-\alpha)S_{xy,l-1}(i) + \alpha\hat{X}_l(i)\hat{Y}_l^*(i); N_u \leqq i \leqq N_o \quad (4$$

$$B_{xy,l}(i) = |S_{xy,l}(i)| \quad (5)$$

Als Glättungskonstante α wird z.B. α = 1 gewählt. Werte $\alpha \ll 1$ sind nicht sinnvoll.

Mit einer Präemphase können höhere Frequenzen angehoben werden. Dies ist dann vorteilhaft, wenn das Sprachsignal und das Störsignal bei höheren Frequenzen eine geringere Leistung aufweisen. Die Werte der Kreuzleistung B_{xy} (i) können z.B. im Bereich 300 bis 1500 Hz um 10dB linear an-steigend angehoben werden. Die Präemphase kann aber auch schon durch die Mikrofoncharakteristik vorgegeben sein.

Aus den Werten B_{xy} (i) werden M Maxima bestimmt und summiert. Es können z.B. M = 8 verwendet werden. Es wird ein aktueller Schätzwert

$$S_{B,l} = \frac{1}{M} \sum^{M} B_{xy,l}(i) \quad (6)$$

bestimmt.

Über einen Impulsmonitor wird eine "simulierte Impulsantwort" S_l berechnet. Das Einschwingverhalten des umgebenden Raumes auf plötzliche energiestarke Schallereignisse (Sprache) wird hiermit grob simuliert (z.B. wird γ = 0.1 gewählt). Die Glättung des Phasenwerts "vom Wortanfang in das Wort hinein" ist mit γ einstellbar.

$$S_{l,1} = (1-\gamma)S_{l,l-1} + \gamma S_{B,l} \quad (7)$$

Außerdem wird über einen Geräuschmonitor eine adaptive Glättungskonstante h berechnet. Mit dieser Glättungskonstanten ergibt sich ein Schätzwert S_N für die Störung. Wurde zuvor eine spektrale Substraktion (SPS) durchgeführt, ist S_N ein Schätzwert für die Reststörung. Für die Glättungskonstante h_o gilt z.B. h_o = 0.03

$$h_l = h_o \frac{2S_{N,l-1}}{S_{N,l-1} + S_{B,l}} \quad (8)$$

$S_{N,L} = (1 - h_l)S_{N,l-1} + h_l S_{B,l}$ (9)

Die Phase der gestörten Signale wird aus den Real- und Imaginärteilen von S_{xy} berechnet. Die Phase wird nur an den M zuvor bestimmten Maxima berechnet.

$$\varphi_1(i) = \arctan \frac{Im[S_{xy,1}(i)]}{Re[S_{xy,1}(i)]} \quad ; \text{für } Re > 0 \qquad (10)$$

und

$$\varphi_1(i) = \pi - \arctan \frac{-Im[S_{xy,1}(i)]}{Re[S_{xy,1}(i)]}; \quad \text{sonst} \qquad (11)$$

Daraus ergibt sich der Phasenanstieg:

$$\varphi'_1(i) = \frac{\varphi_1(i)}{i} \qquad (12)$$

Mit der Länge der Fouriertransformation N und der max. zulässigen Verschiebung um n Taps ergibt sich (N = 256):

$|\phi'|_{max} = |n|\frac{2\pi}{N}$ (13)

Übersteigt der Phasenanstieg $|\phi'|$ an einem der Maxima $|\phi'|_{max}$, so wird dieser Wert ϕ' nicht weiterverwendet. Es wird eine adaptive Glättungskonstante g berechnet:

$$g_1 = \frac{g_o(S_{B,1} - S_{I,1})}{S_{I,1}} \qquad (14)$$

$g_I \leq g_{max}$ (15)

$g_{max} = 0{,}25; g_O = 0{,}25$ (16)

Der aktuelle Wert S_B muß um den Faktor c größer sein als die simulierte Impulsantwort S_I

$S_{B,I} \geq cS_{I,I}$; c = 2 (17)

sonst gilt:

$g_I = 0$ (18)

Der aktuelle Wert S_B muß um den Faktor d größer sein als das Restrauschen S_N

$S_{B,I} \geq dS_{N,I}$; d = 3 (19)

sonst gilt ebenfalls

$g_l = 0$ (20)

Ist Gl. (17) oder Gl. (19) nicht erfüllt, d.h. gilt g = O, so kann die Phasenschätzung abgebrochen werden. Es gilt der alte Phasenschätzwert.

Für alle

$|\phi'_l(i)| \leq |\phi|_{max}$ (21)

gilt:

$$m_{\varphi',l} = \frac{1}{M'} \sum^{M'} \varphi'_l(i) \qquad (22)$$

$$s^2_{\varphi',l} = \frac{1}{M'} \sum^{M'} (\varphi'_l(i))^2 \qquad (23)$$

Von den ursprünglichen M Maxima werden wegen Gl. (21) nur M' für die Gl. (22, 23) verwendet. Ist die Anzahl M' der für die Summen gültigen Werte ϕ kleiner als M_{min}, gilt der geschätzte Phasenanstieg als zu unsicher oder außerhalb des Nutzbereichs (z.B. M_{min} = 6, bei M = 8). Die Phasenschätzung wird dann nicht aktualisiert und das Verfahren hier abgebrochen. Es gilt der alte Phasenschätzwert.

Es wird die Varianz der Schätzung berechnet:

$$\sigma^2_{\varphi',l} = s^2_{\varphi'_l} - m^2_{\varphi'_l} \qquad (24)$$

Als maximale Varianz wird

$\sigma^2_{max} = |\phi'|^2_{max}$ (25)

verwendet.

Entsprechend der Varianz wird die Glättungskonstante g gewichtet. Bei einer großen Streuung gilt:

$$g_l := 0.09 * g_l; \quad \text{für } 0{,}2\sigma^2_{max} < \sigma^2_{\varphi'_l} < \sigma^2_{max} \qquad (26)$$

Bei einer mittleren Streuung gilt:

$$g_l := 0.3 * g_l; \text{ für } 0.02\sigma^2_{max} \leq \sigma^2_{\varphi'_l} \leq 0{,}2\sigma^2_{max} \qquad (27)$$

Bei sehr geringer Streuung gilt:

$$g_l := g_l; \text{ für } \sigma^2_{\varphi'_l} < 0.02\sigma^2_{max} \qquad (28)$$

Entsprechend den Gl. 19 - 22 wird g in der Regel nur am Wortanfang größer Null sein. Dabei muß die Energie des Wortes größer sein als die Energie des Restgeräusches und der simulierten Impulsantwort. Mit der Variablen j wird die aufeinanderfolgende Anzahl für g > 0 gezählt. Entsprechend gilt für die Glättung:

j=1:

$$\varphi'_l = m_{\varphi',l} \tag{29}$$

j=2:

$$m_{\varphi',l} = \frac{(m_{\varphi',l} + m_{\varphi',l-1})}{2} \tag{30}$$

$$\tilde{\varphi}'_l = (1 - 1{,}5 g_l)\ \tilde{\varphi}'_{l-1} + 1{,}5\ g_l m_{\varphi',l} \tag{31}$$

j=3,4...:

$$\tilde{\varphi}'_l = (1 - g_l)\ \tilde{\varphi}'_{l-1} + g_l m_{\varphi',l} \tag{32}$$

Wird z.B. infolge einer Störung die Bedingung g > 0 nur einmal in Folge erfüllt, wird die Phasenschätzung nicht aktualisiert. Eine Aktualisierung der Phasenschätzung erfolgt nur dann, wenn g > 0 mindestens 2-mal in Folge erfüllt wird.

Ein Beispiel für die Zwischengrößen S_B, S_I, S_N' und g und die daraus abgeleitete Phasenschätzung zeigt, FIG. 2. Dabei wird das Wort "Senderwahl" gesprochen und das Fahrgeräusch bei 140km/h addiert. Das Verfahren wird, wie oben angegeben, verwendet. Der Phasenschätzwert ist in Abtastwerten n angegeben. Mit der Größe S_I wird der "Sprachimpuls" teilweise verdeckt und so nur bei starken Energieanstiegen eine Schätzung erlaubt (S_B muß S_I um den Faktor 2 übersteigen). Die Schätzung der Reststörung S_N ermöglicht eine größere Robustheit gegenüber Geräuschen (S_B muß S_N um den Faktor 3 übersteigen).

Patentansprüche

1. Verfahren zur Laufzeitschätzung bei dem Laufzeitunterschiede von geräuschgestörten Signalen von zumindest zwei Sprachkanälen mittels einer Kreuzkorrelation bestimmt werden, dadurch qekennzeichnet,
   - daß im Frequenzbereich die Phasenwerte von zumindest zwei Signalen über eine bestimmte Anzahl von Maxima der Kreuzleistungsdichte ermittelt werden und deren Phasenverschiebung bestimmt wird, und
   - daß der erforderliche Phasenausgleich ebenfalls im Frequenzbereich durchgeführt wird.

2. Verfahren nach Anspruch I, dadurch gekennzeichnet, daß Hintergrundstörungen und das Einschwingverhalten des Raumes bei der Bestimmung der Phasenwerte ständig mitgeschätzt werden.

3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, daß das Hintergrundgeräusch über einen Geräuschmonitor geschätzt wird, und daß ein neuer Phasenwert lediglich dann ermittelt wird, wenn der Schätzwert des Hintergrundgeräusches um einen bestimmten Faktor überschritten wird.

4. Verfahren nach Anspruch 2, dadurch gekennzeichnet, daß das Einschwingverhalten des umgebenden Raumes über einen Impulsmonitor derart geschätzt wird, daß lediglich bei starkem Energieanstieg in den Signalen ein neuer Phasenschätzwert ermittelt wird.

5. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß eine lineare Laufzeitverschiebung der Signale angenommen wird.

6. Verfahren nach einem der vorhergehenden Ansprüche, da-durch gekennzeichnet, daß eine Glättung des Phasenwertes vom Wortanfang in das gesprochene Wort hinein durchgeführt wird, und daß die Varianz der Schätzung bei der Glättung der Phasenwerte mitberücksichtigt wird.

7. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet,
   - daß zumindest zwei Mikrofonsignale x, y mittels einer FFT (Fast Fourier Tansformation) in den Frequenzbereich transformiert werden,
   - daß durch spektrale Substraktion aus den transformierten Signalen die Schätzwerte $\hat{X}$, $\hat{Y}$ bestimmt werden,
   - daß aus den geschätzten Werten $\hat{X}$, $\hat{Y}$ der Betrag der Kreuzleistungsdichte B_{xy} bestimmt wird,
   - daß die Maxima der Kreuzleistungsdichte bestimmt werden, und daß aus einer bestimmten Anzahl Maxima der Kreuzleistungsdichte B_{xy} ein aktueller Wert S_B für die gestörten Signale ermittelt wird, daß abhängig vom aktuellen Wert S_B die Phasen ϕ der gestörten Signale ermittelt werden und damit der Phasenanstieg ϕ' bestimmt wird,
   - daß der Phasenanstieg ϕ'geglättet wird, indem über einen Impulsmonitor ein simulierter Sprachimpuls S_I mit dem aktuellen Wert S_B der gestörten Signale gekoppelt wird, derart, daß eine erneute Phasenschätzung lediglich dann durchgeführt wird, wenn ein starker Energieanstieg des Mikrofonsignals registriert wird, und
   - daß mit einem Geräuschmonitor ein Schätzwert S_N für die Hintergrundgeräuschstörung ermittelt wird und mit dem aktuellen Wert S_B der gestörten Signale gekoppelt wird, derart, daß eine erneute Phasenschätzung lediglich dann durch geführt wird, wenn vom Signal die Hintergrundstörung deutlich überschritten wird.

8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß ein maximaler Phasenanstieg $|\phi'|_{max}$ für die Phase an den einzelnen Maxima vorgebbar ist und eine erneute Phasenschätzung lediglich dann durchgeführt wird, wenn der Phasenanstieg um mindestens M' der M Maxima den maximalen Anstieg $|\phi'|_{max}$ nicht überschreitet.

9. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß die Varianz der Phasenanstiege an den einzelnen Maxima bei der zeitlichen Glättung des Phasenanstiegs berücksichtigt wird.

10. Verfahren nach den Ansprüchen 7 bis 9, dadurch gekennzeichnet, daß eine erneute Phasenschätzung lediglich dann durchgeführt wird, wenn die Bedingungen für einen gültigen Phasenanstieg zeitlich mehrfach in Folge auftreten.

11. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die gestörte Sprache auf mehr als zwei Sprachkanälen aufgenommen wird und daß die Laufzeitunterschiede der einzelnen Kanäle geschätzt werden.

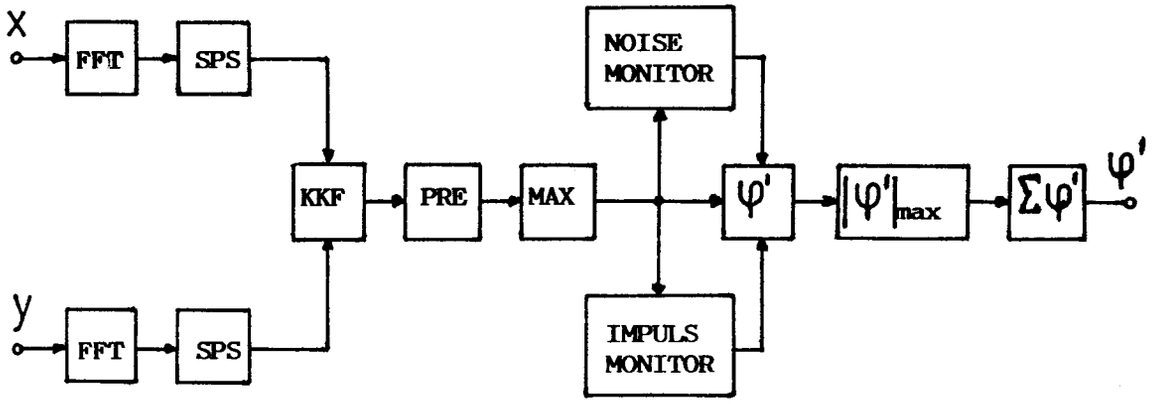
x
FFT
SPS
y
FFT
SPS
KKF
PRE
MAX
NOISE MONITOR
IMPULS MONITOR
φ'
|φ'|max
Σφ'
φ'

FIG.1

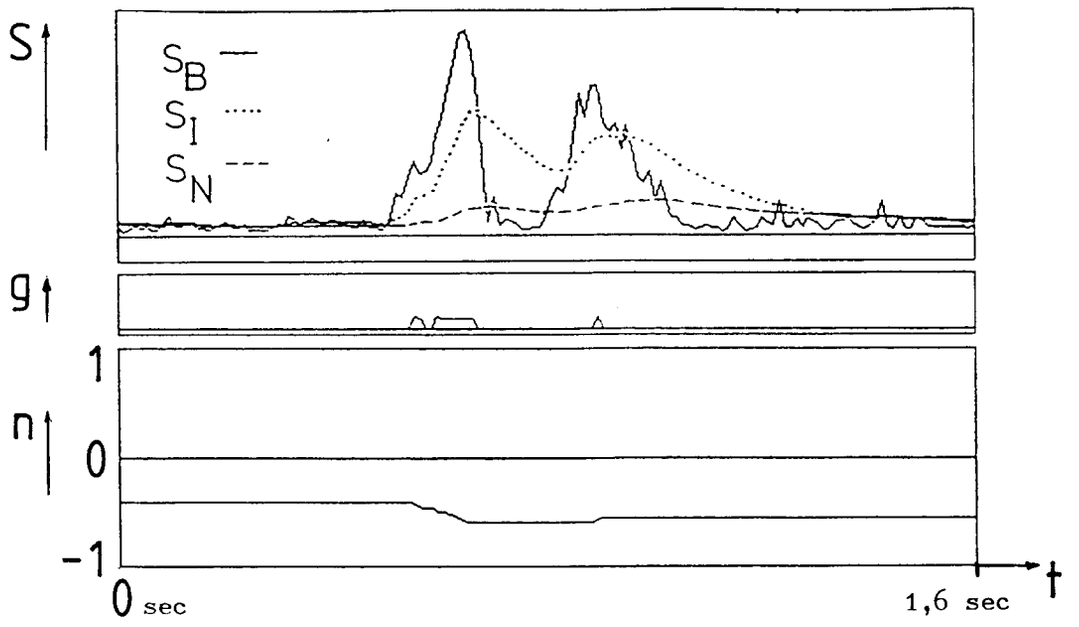
S
SB
SI
SN
g
n
1
0
-1
0 sec
1,6 sec
t

FIG. 2