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

(11) Veröffentlichungsnummer: **0 669 606 A2**

(12) EUROPÄISCHE PATENTANMELDUNG

(21) Anmeldenummer: **95101977.7**

(51) Int. Cl.[6]: **G10L 3/02**

(22) Anmeldetag: **14.02.95**

(30) Priorität: **23.02.94 DE 4405723**

(43) Veröffentlichungstag der Anmeldung:
30.08.95 Patentblatt 95/35

(84) Benannte Vertragsstaaten:
AT CH DE ES FR GB IT LI NL

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

(72) Erfinder: **Linhard, Klaus, Dr.-Ing.**
Gundershofen 86
D-89601 Schelklingen (DE)
Erfinder: **Klemm, Heinz**
Schülinstrasse 31
D-89073 Ulm (DE)

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

(54) Verfahren zur Geräuschreduktion eines gestörten Sprachsignals.

(57) Durch Verfahren zur Geräuschreduktion soll ein gestörtes Sprachsignal möglichst gut von den Störanteilen befreit werden. Das vorliegende Verfahren basiert auf dem Prinzip der spektralen Subtraktion, welche durch eine Medianfilterung ergänzt wird. Die Medianfilterung kann an dem Betragsspektrum des gestörten Eingangssignals oder des Ausgangssignals der spektralen Subtraktion oder an der Übertragungsfunktion der spektralen Subtraktion in Zeitrichtung oder in Frequenzrichtung vorgenommen werden. Die Medianfilterung unterdrückt oder vermeidet insbesondere die als "musical tones" bekannten, bei der spektralen Subtraktion entstehenden Störungen.

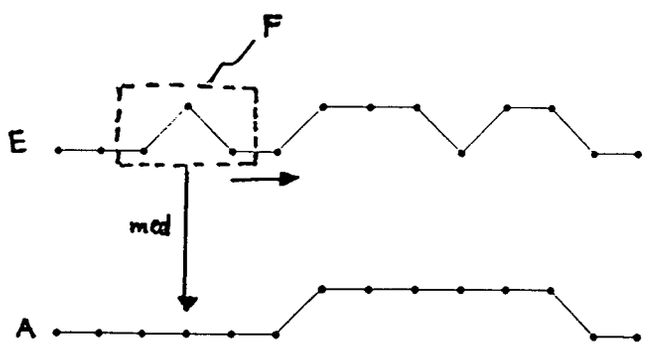

FIG. 1

EP 0 669 606 A2

Die Erfindung betrifft ein Verfahren zur Geräuschreduktion eines gestörten Sprachsignals mit Hilfe der spektralen Subtraktion.

Die Geräuschreduktion mit der Methode der spektralen Subtraktion findet Anwendung bei der automatischen Spracherkennung oder bei Freisprechanlagen zur Verbesserung der Sprachqualität, z.B. beim Telefonieren aus dem Kraftfahrzeug.

Die Geräuschreduktion durch spektrale Subtraktion zeichnet sich dadurch aus, daß relativ stationäre Störungen typischerweise um ca. 10dB reduziert werden können, ohne daß zusätzliche Information über die Störung benötigt wird. Es wird nur der gestörte Sprachkanal benötigt. Das Sprachsignal wird in kurze überlappende Zeitsegmente unterteilt und segmentweise bearbeitet. Bei der spektralen Subtraktion wird in den Sprachpausen ein Schätzwert der Störung ermittelt, und dieser Schätzwert wird im Spektralbereich betragsmäßig subtrahiert. Die spektrale Subtraktion ist auf verschiedene Arten realisierbar, wird aber in der Regel als multiplikatives Filter im Frequenzbereich implementiert. Diese spektrale Subtraktion zeigt den unerwünschten Nebeneffekt eines musikalischen Restgeräuschs, der "musical tones" und einer Sprachverzerrung.

Üblicherweise werden "musical tones" durch eine überhöhte Dämpfung unterdrückt. Die überhöhte Dämpfung kann durch ein Überschätzen der Störung mit einem Überschätzfaktor erfolgen oder durch die Wahl einer speziellen Übertragungskennlinie. Aus der Übertragungskennlinie werden für jede Frequenz die Werte der aktuellen Übertragungsfunktion bestimmt. Es ist üblich im spektralen Subtraktionsfilter eine Betragskennlinie zu implementieren, die eine höhere Dämpfung aufweist als z.B. ein Kennlinie nach dem quadratischen Fehlerkriterium. Speziell entworfene Kennlinien sind ebenfalls möglich. Abhängig von der verwendeten Kennlinie ist eine Überschätzung der Störung um den Faktor 1 bis 3 üblich. Die überhöhte Dämpfung durch die Kennlinie und den Überschätzfaktor ergibt zwar den gewünschten Effekt der Unterdrückung von "musical tones", hat aber auch den Nebeneffekt einer z.T. erheblichen Verzerrung der Sprache.

Eine weitere übliche Methode "musical tones" zu unterdrücken, ist die Maskierung durch Zulassen eines bestimmten Anteils (z.B. 20%) des ursprünglichen Geräuschs als Hintergrundgeräusch ("spectral floor"). "musical tones" werden dadurch weniger hörbar, das Geräusch wird aber auch nicht mehr vollständig unterdrückt.

Es gilt bei der spektralen Subtraktion

$$\hat{S}_{i,l} = K_{i,l} \cdot Y_{i,l} \tag{1}$$

mit

$$Y_{i,l} = S_{i,l} + N_{i,l} \tag{2}$$

und für das Beispiel einer sogenannten Betragskennlinie als Übertragungskennlinie

$$K_{i,l} = 1 - a\sqrt{\frac{|Y_{i,l}|^2}{|\hat{N}_{i,l}|^2}} \tag{3}$$

sowie beispielsweise die Auswahl eines minimalen Übertragungswertes für den spectral floor

$$\mathrm{Min}(K_{i,l}) = b. \tag{4}$$

Mit den Größen:

- $\hat{S}$: geschätztes Ausgangssignal
- K: Übertragungsfunktion
- Y: gestörtes Sprachsignal
- S: Sprachsignal
- N: Störgeräusch
- b: Hintergrundrestgeräusch (spectral floor)
- a: Überschätzfaktor (overestimate)
- $\hat{N}^2$: in Sprachpausen geschätzte Störung
- i: Frequenzindex

l: Zeitindex des Segments

Methoden zur Unterdrückung der "musical tones", durch Kennlinie, "overestimation" und "spectral floor", sind in vielfältiger Variation durch zahlreiche Veröffentlichungen bekannt, z.B.:

Boll, S.: Suppression of Noise in Speech Using the SABER Method, Proc. IEEE Int. Conf. on ASSP, 1978, pp. 600-609.

Boll, S.: Suppression of Acoustic Noise in Speech Using Spectral Substraction, IEEE Trans. on ASSP, Vol. ASSP-27, No. 2, April 79, pp. 113-120.

Berouti, M.; Schwartz, R.; Makhoul, J.: Enhancement of Speech Corrupted by Acoustic Noise, Proc. Int. Conf. on ASSP, 1979, pp. 208-211.

Vary, P.: Noise Suppression by Spectral Magnitude Estimation - Mechanism and Theoretical Limits-, Signal Processing, Vol. 8, No. 4, 1986, pp. 387-400.

Xie, F.; Compernolle.: Speech Enhancement by Nonlinear Spectral Estimation - A Unifying Approach, Int. Conf. Eurospeech, 1993, pp. 617-620.

Über die oben angesprochenen Methoden hinaus, sind weitere spezielle Methoden bekannt, die ebenfalls zur Reduzierung der "musical tones" verwendet werden:

Die Amplitudenwerte zeitlich aufeinanderfolgender gestörter Sprachspektren werden gemittelt (z.B. bei Boll "magnitude averaging"). Dadurch werden zwar Rauschanteile gedämpft aber da Sprache stark instationär ist, tritt schon bei kurzen Mittelungslängen eine zeitliche Verschmierung des Sprachsignals auf (echoartiger Effekt). Bei Boll wird weiterhin ein "magnitude plus bandwith measurement test " beschrieben, nachdem spektrale Bereiche mit einer Bandbreite unter 300Hz und einer Amplitude, kleiner als eine vorgegebene Schwelle, als "residual noise" erkannt werden. Diese Bereiche werden dann zusätzlich gedämpft. Es wird von Boll vorgeschlagen, den "residual noise" dadurch zu reduzieren, daß aus drei zeitlich aufeinanderfolgenden Spektren des gefilterten Signals jeweils der minimale Wert als Ausgangssignal verwendet wird. Die Ausgabe der minimalen Spektrallinie von drei zeitlich benachbarten Linien führt zwar zu einer deutlichen Reduzierung des Restgeräuschs und damit der "musical tones", gelegentlich treten jedoch in unregelmäßigen Abständen plötzliche kurze "Geräuschbündel" auf.

Ein weiteres Verfahren verwendet eine sogenannte nichtlineare spektrale Subtraktion. Der Überschätzfaktor wird hier abhängig vom Pausengeräusch und dem aktuell anliegenden Signal errechnet. Die optimale Einstellung dieser Regelung ist jedoch schwierig. (Lockwood, P.; Boudy, J.: Experiments with a Nonlinear Spectral Subtraction (NSS), Hidden Markov Models and the projection, for robust speech recognition in cars, Speech Communication, No. 11, 1992, p. 215-228).

Aufgabe der vorliegenden Erfindung ist es, ein Verfahren zur Geräuschreduktion eines gestörten Sprachsignals anzugeben, welches bei hoher Sprachqualität des Ausgangssignals eine starke Reduktion der Geräusche, insbesondere auch der "musical tones" ermöglicht.

Erfindungsgemäße Lösungen dieser Aufgabe sowie vorteilhafte Ausführungen und Weiterbildungen sind in den Patentansprüchen beschrieben.

Die Medianfilterung erweist sich als vorteilhaftes Verfahren zur weiteren wesentlichen Verbesserung des Verfahrens der spektralen Subtraktion für die Geräuschreduktion eines gestörten Sprachsignals. Die Medianfilterung kann dabei sowohl auf das Betragsspektrum des gestörten Eingangssignals oder des nach der spektralen Subtraktion geräuschreduzierten Ausgangssignals als auch auf die aus der Anwendung einer Übertragungskennlinie bestimmten Übertragungsfunktion angewandt und in Zeitrichtung oder in Frequenzrichtung durchgeführt werden. Das Betragsspektrum des Sprachsignals setzt sich entsprechend der Segmentierung des Sprachzeitsignals aus einer Folge von Segmentspektren zusammen. Die Übertragungsfunktion ist repräsentiert durch die zeit- und frequenzdiskreten Werte $K_{i,l}$ (z.B. Gleichung (3)). Auch eine Kombination verschiedener dieser Vorgehensweisen kann vorteilhaft sein. So sieht ein bevorzugtes Verfahren vor, in Sprachpausen durch Anwendung der Medianfilterung bevorzugt in zeitlicher Richtung, auf die Übertragungsfunktion den natürlichen Eindruck eines schwachen Hintergrundgeräusches zu bewahren und während Sprachaktivität durch Anwendung der Medianfilterung auf das Betragsspektrum des Sprachsignals eine starke Unterdrückung der "musical tones" zu erreichen. Die getrennte Erkennung von Sprachpausen und Sprachaktivität ist zur Ermittlung eines mittleren Geräuschsignals während Sprachpausen ohnehin vorgesehen und bekannt, so daß hierfür kein gesonderter Aufwand erforderlich ist. Die erfindungsgemäßen Verfahren sind einfach implementierbar.

Das Prinzip der Medianfilterung an sich ist allgemein bekannt (z.B. Mitra, S.K.: Handbook for Digital Signal Processing, John Wiley & Sons, 1993).

Die Medianfilterung ist auch zur Verarbeitung von Sprachsignalen bereits bekannt. So wird beispielsweise in DE 32 43 231 A1 und DE 32 43 232 A1 ein Medianfilter auf aufeinanderfolgende Kurzzeitmittelwerte, die ein Maß für die mittlere Leistung von Sprachsignalabschnitten darstellen, als Glättungsfilter angewandt. Durch Vergleich der geglätteten Wertefolge mit einem Schwellwert werden Sprachpausen erkannt. Eine

Störbefreiung des Sprachsignals findet dadurch nicht statt.

In IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-23, No. 6, Dec. 1975, S. 552-557 ist die Anwendung eines Medianfilters im Kombination mit einem linearen Glättungsfilter auf Abtastwerte der Intensität eines Sprachsignals beschrieben. Eine Signalverarbeitung im Spektralbereich ist nicht vorgesehen und es kann nur eine geringe Störüberlagerung bewältigt werden.

FIG. 1 zeigt ein Beispiel für ein Eingangssignal E und ein mit einem Medianfilter der Länge 3 gefiltertes Ausgangssignal A. Das Medianfilter sortiert zuerst die Werte innerhalb des Datenfensters F und gibt dann den mittleren Wert med aus. Das Medianfilter blendet kurze Signalspitzen aus, erhält aber die übrigen Signalflanken.

Für das Beispiel der Anwendung eines Medianfilters der Länge 3 auf ein geräuschreduziertes Betragsspektrum eines Sprachsignals gilt bei zeitlicher Filterung

$$|\hat{S}^{m}_{i,l}| = med(\,|\hat{S}_{i,l-1}|,\ |\hat{S}_{i,l}|,\ |\hat{S}_{i,l+1}|\,) \qquad (5)$$

oder bei Filterung in Frequenzrichtung

$$|\hat{S}^{m}_{i,l}| = med(\,|\hat{S}_{i-1,l}|,\ |\hat{S}_{i,l}|,\ |\hat{S}_{i+1,l}|\,) \qquad (6)$$

Der Filterung am Betrag ist die Filterung am Betragsquadrat im Prinzip gleichwertig.

Die Wirkung der Medianfilterung auf die Verringerung der "musical tones" ist veranschaulicht anhand von Darstellungen eines typischen zeitlichen und spektralen Verlaufs solcher "musical tones". Dargestellt ist das Betragsspektrum eines in einer Sprachpause gewonnenen und mit Hilfe der spektralen Subtraktion geschätzten Ausgangssignals. Da in der Sprachpause keine Sprachanteile vorliegen treten vor allem die "musical tones" deutlich in Erscheinung.

Als Beispiel der spektralen Subtraktion wurde verwendet: Standardverfahren mit Betragskennlinie, 20% Hintergrundgeräusch (b = 0,2), ohne Überschätzfaktor (a = 1,0).

Als Geräusch-Beispiel wurde verwendet: Fahrzeuginnengeräusch bei 140km/h, 12kHz Abtastfrequenz, Segmentlänge 512 Werte, die letzten 256 Werte jedes Segments werden zu Null gesetzt, die ersten 256 Werte jedes Segments werden mit Hanning-Fenster multipliziert, Segmente sind halb überlappt, d.h. alle 10,67ms ein neues Segment.

FIG. 2 zeigt zunächst über der Frequenz (linear 0 bis 6kHz) das Spektrum für 4 zeitlich aufeinanderfolgende Segmente (Zeitabstand 10,67ms, Index l) und dann über der Zeit (0 bis 2,5sec) den Signalverlauf für 4 aufeinanderfolgende diskrete Frequenzen (Index i), stellvertretend für alle 256 Frequenzen. Es zeigt sich als typische Eigenschaft der "musical tones", daß der Verlauf über der Frequenz relativ ausgedehnte Störungen (breite Impulse) aufweist, wogegen der Verlauf über der Zeit einen starken impulsartigen Charakter (schmale Impulse) hat. Genau der impulsartige Charakter in zeitlicher Richtung macht die Medianfilterung hier besonders effektiv. Eine impulsartige Störung wird gelöscht. Für impulsartige Störungen mit breiteren Impulsen ist eine größere Fensterlänge des Medianfilters erforderlich. Im Gegensatz zu linearen Filterungsverfahren (Glättungsfilter, "linear smoother") findet keine Verschmierung des Signalverlaufs statt. Die Darstellung der in zeitlicher Richtung mit dem 3-er Median gefilterten Signale in FIG. 3 verdeutlicht diese Eigenschaft. Das gefilterte Signal zeigt im Zeitverlauf deutlich einen glatteren Verlauf. Im Frequenzverlauf sind einige der (breiteren) Impulse durch die Filterung in Zeitrichtung ebenfalls gelöscht.

Bei Sprachaktivität führt die Anwendung des Medianfilters in zeitlicher Richtung der einzelnen Spektrallinien zu einer Verbesserung der Sprachqualität, da impulsartige Störungen des Sprachspektrums "repariert" werden. Das Sprachsignal selbst wird nur sehr gering verändert. Eine Erhöhung der Fensterlänge von 3 auf 5 (in Zeitrichtung) ergibt zwar eine noch bessere Auslöschung der "musical tones", es wird aber bereits ein schwacher echoartiger Charakter der Sprache hörbar.

Das Medianfilter kann anstatt am Ausgangssignal auch am Eingangssignal, vor der spektralen Subtraktion, durchgeführt werden. Im Idealfall können dadurch keine "musical tones" entstehen, die sonst alternativ durch die Nachfilterung mit dem Medianfilter gelöst werden. Die Medianfilterung am Eingangssignal kann dann vorteilhaft sein, wenn "musical tones" die verschiedenen implementierten Verarbeitungsschritte im spektralen Substraktionsfilter (außer der Kennlinienfunktion) beeinflussen. Es soll im weiteren nicht auf

mögliche Vor- oder Nachteile einer Medianfilterung am Ein- oder Ausgangssignal eingegangen werden. Im Prinzip sind beide Möglichkeiten gegeben und von speziellen Fällen der Implementierung abgesehen gleichwertig.

Das Medianfilter kann anstatt am Betragsspektrum eines Sprachsignals auch an der Übertragungsfunktion K ausgeführt werden.

Es gilt für den 3-er Median:

$$|K_{i,l}^{m}| = med(\,|K_{i,l-1}|,\ |K_{i,l}|,\ |K_{i,l+1}|\,) \qquad (7)$$

oder

$$|K_{i,l}^{m}| = med(\,|K_{i-1,l}|,\ |K_{i,l}|,\ |K_{i+1,l}|\,) \qquad (8)$$

FIG. 4 zeigt die Übertragungsfunktion K über der Zeit und über der Frequenz. Dargestellt ist der gleiche Ausschnitt wie in FIG. 2. Die Übertragungsfunktion zeigt ein ähnliches Verhalten wie das Ausgangssignal in FIG. 2.

FIG. 5 zeigt die in zeitlicher Richtung mit dem 3-er Median gefilterte Übertragungsfunktion. Dargestellt ist der gleiche Ausschnitt wie in FIG. 3. Auch hier ist die Medianfilterung in zeitlicher Richtung aus den gleichen Gründen wie beim Ausgangssignal äußerst effektiv.

Die effektive Unterdrückung der "musical tones" durch die Medianfilterung kann wie folgt erklärt werden:
Ein Eingangssignal mit einer impulsartigen Störung verursacht die entsprechende impulsartige Änderung der Übertragungsfunktion. Im ursprünglichen Geräusch gehört dieser lokale Impuls zum natürlichen Geräusch und wird deshalb nicht als besonders störend empfunden. Das Spektrum des Eingangssignals wird mit der Übertragungsfunktion multipliziert. Die impulsartige Störung wird dadurch zusätzlich verstärkt ist jetzt als "musical tone" hörbar.

Die impulsunterdrückende Eigenschaft der Medianfilterung wirkt sich besonders deutlich auf die verstärkte Impulsstörung und somit auf die "musical tones" aus. Die Medianfilterung wirkt reparierend auf die impulsartige Störung.

Die Medianfilterung am Betragsspektrum des Eingangs- oder Ausgangssignals ergibt gegenüber der Medianfilterung an den Übertragungswerten den höheren Gewinn an der Unterdrückung von impulsartigen Störungen, kann aber auch zu besonders in Sprachpausen auffallenden als unnatürlich empfundenen Veränderungen führen, während die Medianfilterung der Übertragungswerte in Sprachpausen im wesentlichen zu einer reinen Dämpfung des Signals führt, das dadurch leiser aber natürlich klingt. Im Idealfall entstehen keine "musical tones". Eine bevorzugte Ausführungsform der Erfindung macht sich dies zunutze, indem die Medianfilterung bei Sprachaktivität am Betragsspektrum und in Sprachpausen an den Übertragungswerten durchgeführt wird. Die erforderliche Sprach-Pausen-Entscheidung steht bei der spektralen Subtraktion ohnehin zur Verfügung, da die Bildung des Geräuschschätzwertes nur in den Sprachpausen durchgeführt wird.

Anstelle der Medianfilterung in Zeitrichtung wie beschrieben kann auch eine Medianfilterung in Frequenzrichtung gemäß Gleichung (6) durchgeführt werden. Die gegebenen ausführlichen Darlegungen gelten für die Filterung in Frequenzrichtung analog. Es zeigt sich, daS mit abnehmender Zahl der Abtastwerte innerhalb eines Zeitsegments die Medianfilterung in Frequenzrichtung an Vorteilen gewinnt gegenüber der Filterung in Zeitrichtung und umgekehrt.

Bei den gebräuchlichen Werten für die Segmentlänge nach Zeit- und Abtastwerten ist die Anwendung der Medianfilterung in Zeitrichtung besonders vorteilhaft.

Bei der beschriebenen Anwendung einer Medianfilterung in zeitlicher Richtung mit den beispielhaft angegebenen Werten für Abtastrate und Fensterlänge ist die Fensterlänge wie im Beispiel angegeben gleich der minimalen Medianfensterlänge 3. Größere Fensterlängen führen in diesem Falle zwar zu einer weiteren Unterdrückung der "musical tones", u.U. aber auch zu einer als unnatürlich empfundenen Einebnung des Sprachsignals. Die bevorzugte Fensterlänge ist daher 3 wie beispielhaft angegeben. Für zeitlich kürzere Segmente kann eine größere Fensterlänge bei der Medianfilterung angemessen sein. Der von dem Fenster der zeitlichen Medianfilterung abgedeckte Zeitintervall sollte aber 50ms nicht überschrei-

ten.

Für die Filterung in Frequenzrichtung orientiert sich die Fensterlänge des Medianfilters an der Datensegmentlänge. Die Datensegmentlänge sollte im zahlenmäßig beschriebenen Beispiel kleiner als 64 sein, das Medianfilter nicht größer als 5.

Patentansprüche

1. Verfahren zur Geräuschreduktion eines gestörten Sprachsignals mit Hilfe der spektralen Subtraktion, dadurch gekennzeichnet, daß das Betragsspektrum des Sprachsignals einer Medianfilterung unterzogen wird.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die Medianfilterung auf das Betragsspektrum des gestörten Eingangssignals angewandt wird.

3. Verfahren nach einem der Ansprüche 1 und 2, dadurch gekennzeichnet, daß die Medianfilterung auf das Betragsspektrum des Ausgangssignals der spektralen Subtraktion angewandt wird.

4. Verfahren zur Geräuschreduktion eines gestörten Sprachsignals mit Hilfe der spektralen Subtraktion, wobei aus einer vorgebbaren Übertragungskennlinie eine Übertragungsfunktion für die spektrale Subtraktion bestimmt wird, dadurch gekennzeichnet, daß die Übertragungsfunktion einer Medianfilterung unterzogen werden.

5. Verfahren zur Geräuschreduktion eines gestörten Sprachsignals mit einer Kombination vorhergehender Ansprüche.

6. Verfahren nach Anspruch 5, dadurch gekennzeichnet, daß die Medianfilterung in Sprachpausen auf die Übertragungswerte und bei Sprachaktivität auf das Betragsspektrum des Sprachsignals angewandt wird.

7. Verfahren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, daß die Medianfilterung in zeitlicher Richtung angewandt wird.

8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß die Fensterlänge des Medianfilters drei aufeinanderfolgende Zeitsegmente umfaßt.

9. Verfahren nach Anspruch 7 oder Anspruch 8, dadurch gekennzeichnet, daß die Fensterlänge des Medianfilters kleiner als 50ms ist.

10. Verfahren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, daß die Medianfilterung in Frequenzrichtung angewandt wird.

11. Verfahren nach Anspruch 10, dadurch gekennzeichnet, daß die Fensterlänge des Medianfilters nicht mehr als 5 Frequenzwerte umfaßt.

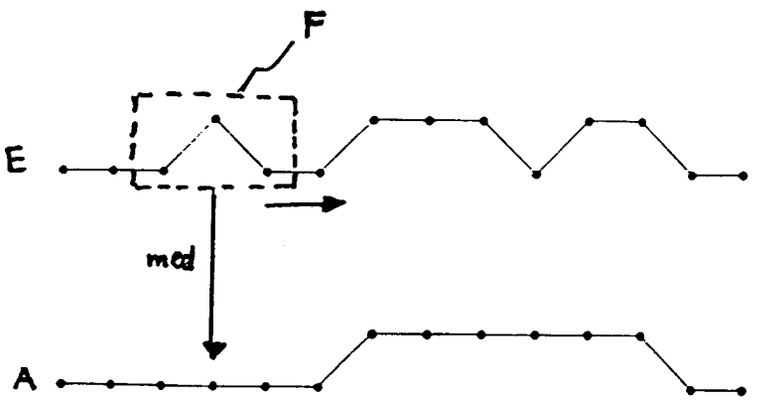

FIG. 1

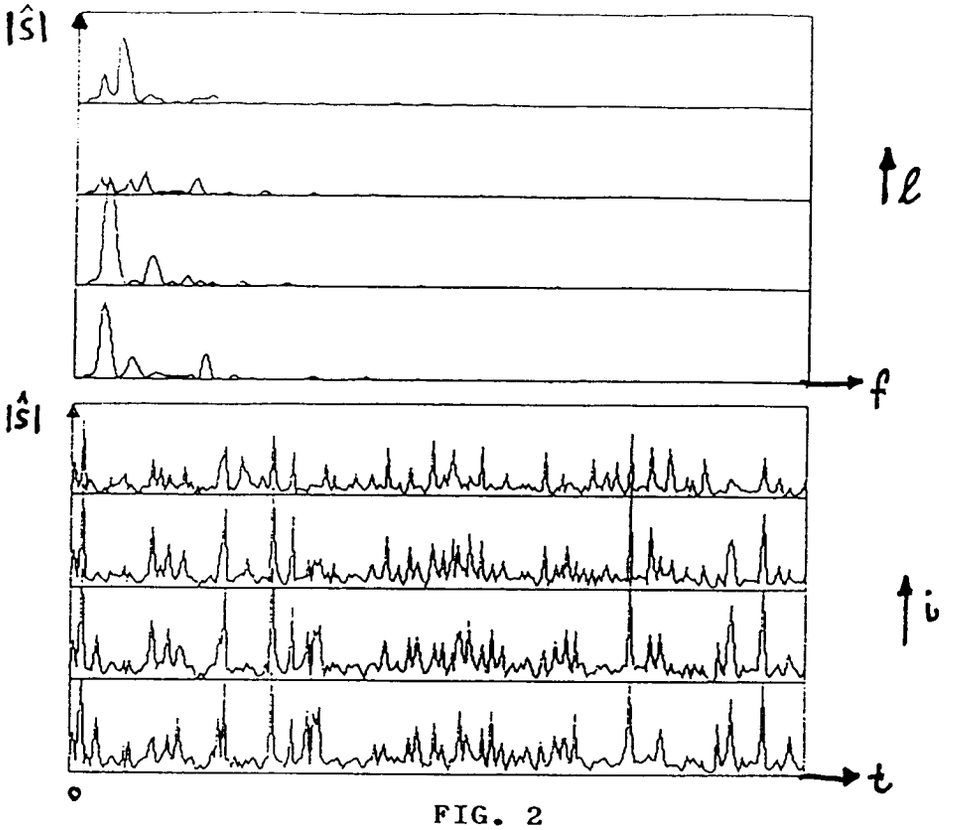

FIG. 2

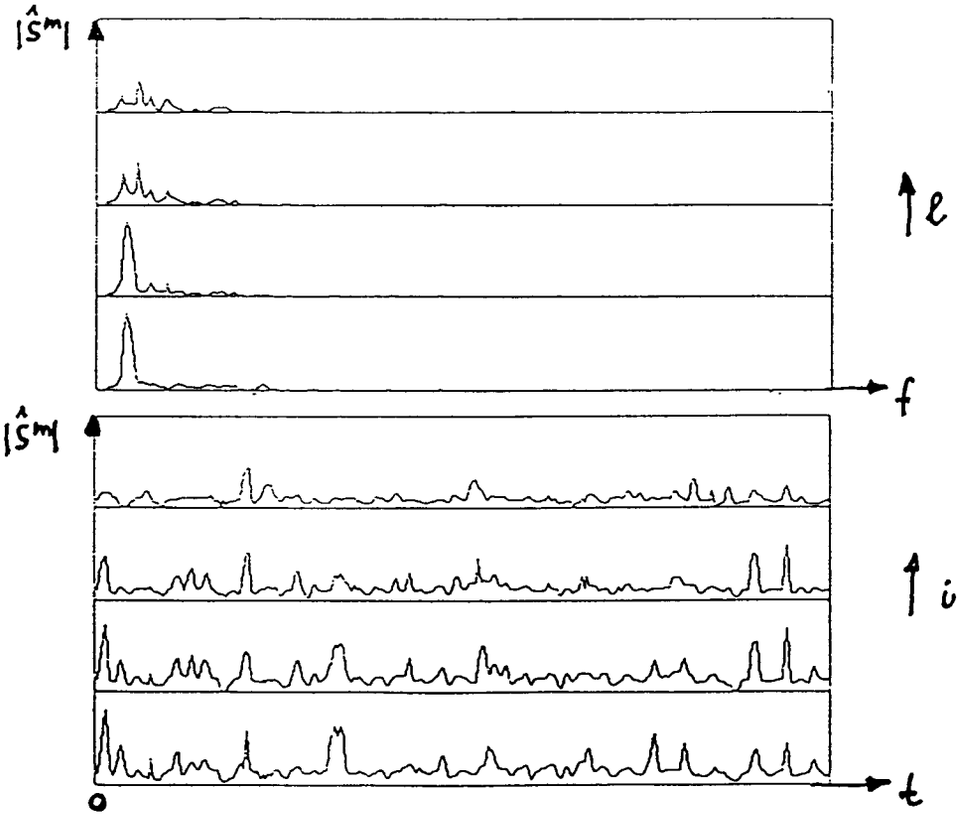

FIG. 3

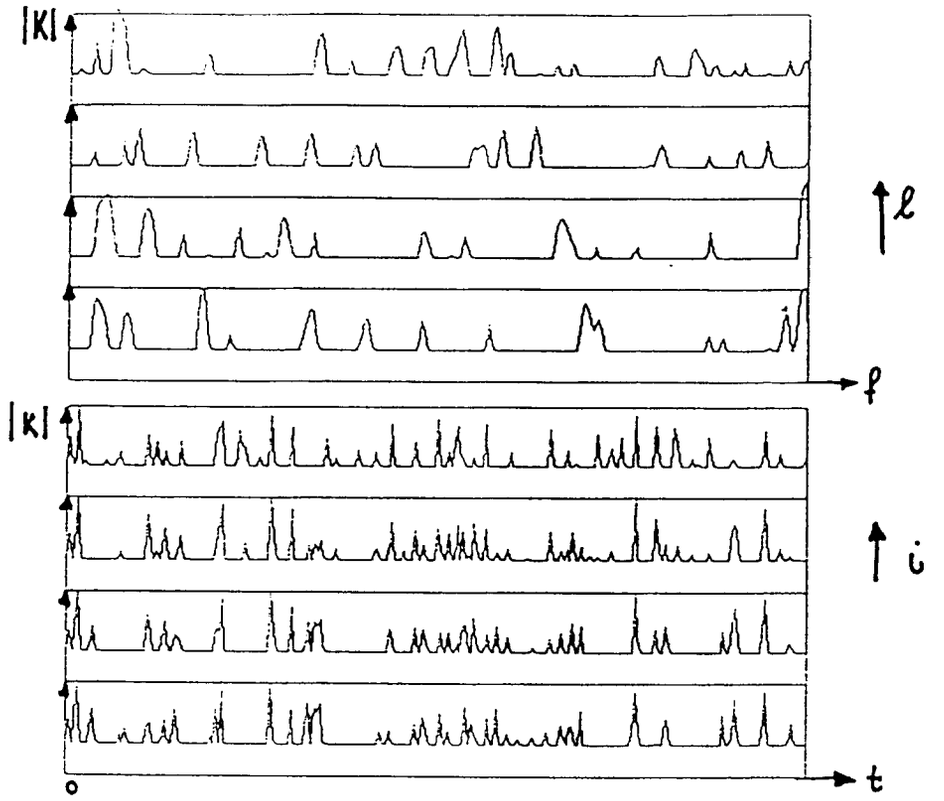

FIG. 4

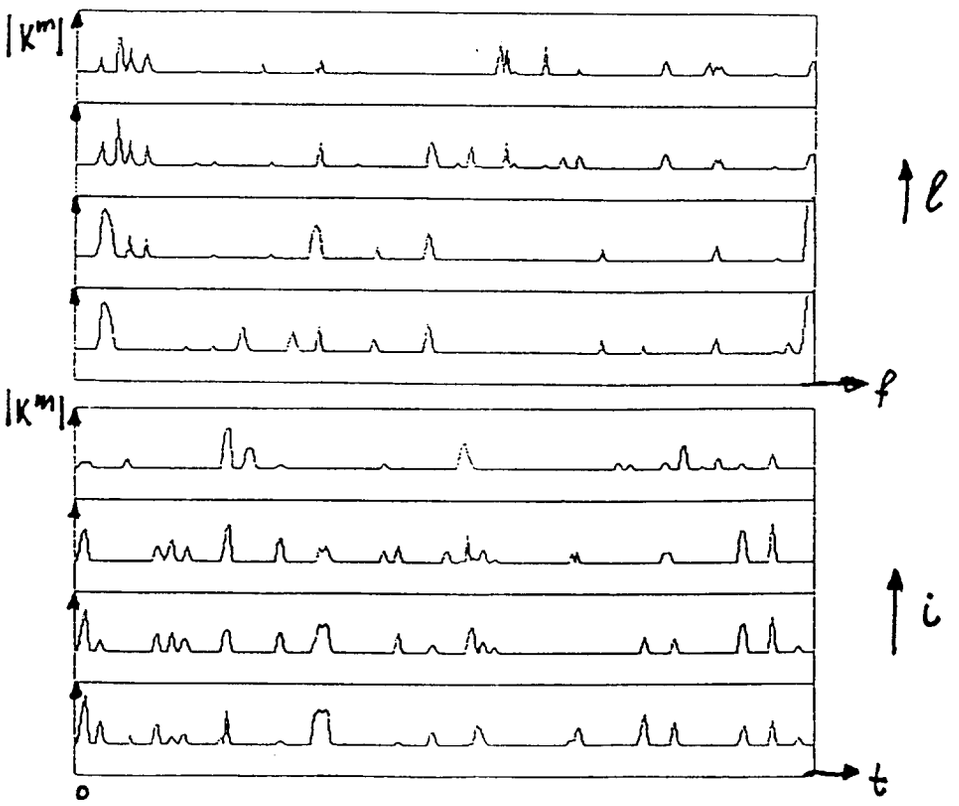

FIG. 5