



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) **EP 1 014 340 A2**

(12) **EUROPÄISCHE PATENTANMELDUNG**

(43) Veröffentlichungstag:  
**28.06.2000 Patentblatt 2000/26**

(51) Int. Cl.<sup>7</sup>: **G10L 21/02**

(21) Anmeldenummer: **99125575.3**

(22) Anmeldetag: **21.12.1999**

(84) Benannte Vertragsstaaten:  
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE**  
Benannte Erstreckungsstaaten:  
**AL LT LV MK RO SI**

(30) Priorität: **21.12.1998 DE 19859174**

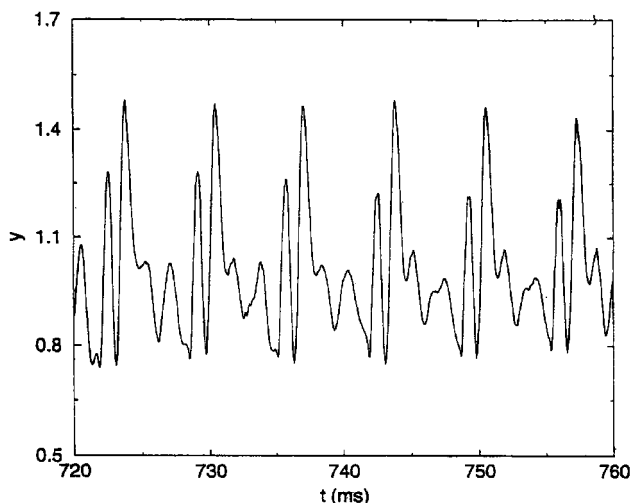
(71) Anmelder:  
**Max-Planck-Gesellschaft zur Förderung  
der Wissenschaften e.V.  
80539 München (DE)**

(72) Erfinder:  
• **Hegger, Rainer, Dr.  
01159 Dresden (DE)**  
• **Kantz, Holger, Dr.  
01187 Dresden (DE)**  
• **Matassini, Lorenzo  
01187 Dresden (DE)**

(74) Vertreter: **Hertz, Oliver, Dr.  
v. Bezold & Partner,  
Patentanwälte  
Akademiestrasse 7  
80799 München (DE)**

(54) **Verfahren und Vorrichtung zur Verarbeitung rauschbehafteter Schallsignale**

(57) Zur Signalverarbeitung eines Schallsignals  $y$ , in dem eine Redundanz enthalten ist, die hauptsächlich aus annähernden Wiederholungen von Signalprofilen besteht, werden die Signalprofile erfasst und Korrelationen zwischen den Signalprofilen innerhalb von Abschnitten des Schallsignals erfasst. Korrelierte Signalanteile werden einem Leistungsanteil und unkorrelierte Signalanteile einem Rauschanteil des Schallsignals zugeordnet. Die Korrelationen zwischen den Signalprofilen werden mit Verfahren der nichtlinearen Rauschreduzierung in deterministischen Systemen in rekonstruierten Vektorräumen ermittelt, die auf der Zeitdomäne basieren.



**Fig. 2**

**EP 1 014 340 A2**

## Beschreibung

**[0001]** Die Erfindung betrifft Verfahren zur Verarbeitung rauschbehafteter Schallsignale, insbesondere zur nichtlinearen Rauschreduzierung in Sprachsignalen, zur nichtlinearen Trennung von Leistungs- und Rauschsignalen und zur Anwendung nichtlinearer Zeitreihenanalysen, die auf dem Konzept des niedrigdimensionalen deterministischen Chaos beruhen. Die Erfindung betrifft auch eine Vorrichtung zur Implementierung der Verfahren und deren Verwendung.

**[0002]** Die Rauschreduzierung bei der Aufnahme, Speicherung, Übertragung oder Wiedergabe menschlicher Sprache besitzt eine hohe technische Relevanz. Rauschen kann als reine Meßungenauigkeit z.B. in Form des Digitalfehlers bei Ausgabe von Schallamplituden, als Rauschen im Übertragungskanal oder als dynamisches Rauschen durch die Kopplung des betrachteten Systems mit der Außenwelt auftreten. Beispiele für Rauschreduzierungen der menschlichen Sprache sind allgemein aus der Telekommunikation, der automatischen Spracherkennung oder dem Einsatz elektronischer Hörhilfen bekannt. Das Problem der Rauschreduzierung tritt nicht nur bei menschlicher Sprache, sondern auch bei anderen Arten von Schallsignalen, und nicht nur bei stochastischem Rauschen, sondern auch bei allen Formen der Überlagerung eines relevanten Schallsignals durch Fremdgeräusche auf. Es besteht ein Interesse an einem Signalverarbeitungsverfahren, mit dem stark aperiodische und nicht-stationäre Schallsignale in Bezug auf Leistungs- und Rauschanteile analysiert, manipuliert oder getrennt werden können.

**[0003]** Ein typischer Ansatz zur Rauschreduzierung, d.h. zur Zerlegung eines Signals in bestimmte Leistungs- und Rauschanteile, beruhen auf einer Signalfilterung im Frequenzbereich. Im einfachsten Fall erfolgt die Filterung mit Bandpaßfiltern, woraus jedoch das folgende Problem entsteht. In der Regel ist stochastisches Rauschen breitbandig (häufig sogenanntes „weißes Rauschen“). Wenn das Leistungssignal jedoch selbst stark aperiodisch und somit breitbandig ist, wird mit dem Frequenzfilter auch ein Leistungssignalanteil zerstört, woraus sich unzulängliche Ergebnisse ergeben. Soll z.B. bei einer Sprachübertragung die menschliche Sprache durch einen Tiefpaßfilter von hochfrequentem Rauschen befreit werden, so wird das Sprachsignal verzerrt.

**[0004]** Ein weiterer allgemein bekannter Ansatz zur Rauschreduzierung besteht in der Rauschkompensation bei Schallaufnahmen. Dabei wird beispielsweise mit einem ersten Mikrofon die von einem Lärmpegel in einem Raum überlagerte menschliche Sprache und mit einem zweiten Mikrofon ein Schallsignal aufgenommen, das im wesentlichen den Lärmpegel repräsentiert. Vom Meßsignal des zweiten Mikrofons wird ein Kompensationssignal abgeleitet, das bei Überlagerung mit dem Meßsignal des ersten Mikrofons das Rauschen aus

dem umgebenden Raum kompensiert. Diese Technik ist aufgrund des relativ hohen Geräteaufwandes (Einsatz spezieller Mikrofone mit Richtcharakteristik) und wegen des eingeschränkten Einsatzbereiches z.B. bei der Sprachaufnahme nachteilig.

**[0005]** Es sind ferner Verfahren zur nichtlinearen Zeitreihenanalyse auf der Grundlage des Konzepts des niedrigdimensionalen deterministischen Chaos bekannt. Da in fast allen Bereichen unserer täglichen Umwelt, aber auch in vielen Bereichen der Wissenschaft und Technik komplexes dynamisches Verhalten eine wichtige Rolle spielt, z.B. wenn Vorgänge in der Medizin, Ökonomie, Signaltechnik oder Meteorologie aperiodische, schwer vorhersagbare und oft auch schwer klassifizierbare Signale liefern, stellt die Zeitreihenanalyse einen grundsätzlichen Ansatz dar, aus beobachteten Daten möglichst viel über die Eigenschaften oder den Zustand eines Systems zu lernen. Bekannte Analyseverfahren zum Verständnis aperiodischer Signale werden z.B. von H. Kantz et al. in „Nonlinear Time Series Analysis“, Cambridge University Press, Cambridge, 1997, oder von H. D. I. Abarbanel in „Analysis of Observed Chaotic Data“, Springer, New York, 1996 beschrieben. Diese Verfahren basieren auf dem Konzept des deterministischen Chaos. Deterministisches Chaos bedeutet, daß zwar ein Systemzustand zu einem bestimmten Zeitpunkt in eindeutiger Weise den Systemzustand zu jedem beliebigen späteren Zeitpunkt festlegt, das System aber dennoch über längere Zeit unvorhersagbar ist. Dies ergibt sich daraus, daß der aktuelle Systemzustand mit einem unvermeidlichen Fehler erfaßt wird, dessen Wirkung je nach Bewegungsgleichung des Systems exponentiell wächst, so daß nach relativ kurzer Zeit ein simulierter Modellzustand mit dem Realzustand des Systems keine Ähnlichkeit mehr besitzt.

**[0006]** Für Zeitreihen deterministischer chaotischer Systems wurden Verfahren zur Rauschunterdrückung entwickelt, die keine Trennung im Frequenzbereich vornehmen, sondern explizit auf die deterministische Struktur des Signals zurückgreifen. Diese Verfahren werden beispielsweise von P. Grassberger et al. in „CHAOS“, Bd. 3, 1993, S. 127, von H. Kantz et al. (s.o.) und von E.J. Kostelich et al. in „Phys. Rev. E“, Bd. 48, 1993, S. 1752, beschrieben. Das Prinzip der Rauschunterdrückung für deterministische Systeme wird im folgenden unter Bezug auf Fig. 10a beschrieben.

**[0007]** Fig. 10 zeigt schematisch die Abhängigkeit aufeinanderfolgender Zeitreihenwerte für rauschfreie bzw. verrauschte Systeme (am Beispiel eines eindimensionalen Zusammenhangs). Die rauschfreien Daten eines deterministischen Systems liefern das in Fig. 10a gezeigte Bild. Es besteht eine exakte (hier: eindimensionale) deterministische Beziehung zwischen einem Wert und dem Folgewert. Die Zeitversatzvektoren, zu denen Einzelheiten weiter unten erläutert werden, liegen in einer niedrigdimensionalen Mannigfaltigkeit im Einbettungsraum. Bei Einführung von Rau-

schen wird die deterministische Beziehung durch eine approximative Beziehung ersetzt. Die Daten liegen nicht mehr auf der Untermannigfaltigkeit, sondern in ihrer Nähe (Fig. 10b). Die Unterscheidung von Leistung und Rauschen erfolgt über die Dimensionalität. Alles, was aus der Untermannigfaltigkeit herausführt, ist auf den Einfluß des Rauschens zurückzuführen.

**[0008]** Dementsprechend erfolgt die Rauschunterdrückung für deterministisch chaotische Signale in drei Schritten. Zuerst wird die Dimension  $m$  des Einbettungsraumes und die Dimension der Mannigfaltigkeit, in der die unverrauschten Daten liegen, abgeschätzt. Zur eigentlichen Korrektur wird dann für jeden einzelnen Punkt die Mannigfaltigkeit in seiner Nähe identifiziert und schließlich zur Rauschreduzierung der betrachtete Punkt auf die Mannigfaltigkeit projiziert (Fig. 10c).

**[0009]** Der Nachteil der illustrierten Rauschunterdrückung besteht in deren Beschränkung auf deterministische Systeme. In einem nicht-deterministischen System, in dem also kein eindeutiger Zusammenhang zwischen einem Zustand und einem Folgezustand besteht, ist das Konzept der Identifizierung einer Bewegungsgleichung und der Betrachtung einer glatten Mannigfaltigkeit, wie es in Fig. 10 illustriert ist, nicht anwendbar. So bilden beispielsweise die Signalamplituden von Sprachsignalen Zeitreihen, die unvorhersagbar sind und den Zeitreihen nicht-deterministischer Systeme entsprechen.

**[0010]** Die Anwendbarkeit der herkömmlichen nichtlinearen Rauschreduzierung ist für Sprachsignale insbesondere aus den folgenden Gründen bisher ausgeschlossen. Die menschliche Sprache (aber auch andere Schallsignale natürlichen oder synthetischen Ursprungs) ist in der Regel hochgradig nichtstationär. Die Sprache ist aus einer Verkettung von Phonemen zusammengesetzt. Die Phoneme wechseln sich ständig ab, so daß sich die Sprachdynamik laufend ändert. Beispielsweise enthalten Zisch-Laute vorrangig hohe Frequenzen und Vokale (z.B. „o“) vorrangig niedrige Frequenzen. Damit wären zur Beschreibung der Sprache Bewegungsgleichungen erforderlich, die sich ständig im Zeitverlauf ändern. Die Existenz einer einheitlichen Bewegungsgleichung ist jedoch Voraussetzung für das unter Bezug auf Fig. 10 beschriebene Konzept der Rauschunterdrückung.

**[0011]** Es ist die Aufgabe der Erfindung, ein verbessertes Signalverarbeitungsverfahren für Schallsignale, insbesondere für verrauschte Sprachsignale, anzugeben, mit dem eine effektive und schnelle Trennung von Leistungs- und Rauschanteilen des betrachteten Schallsignals möglichst verzerrungsfrei möglich ist. Die Aufgabe der Erfindung ist es auch, Vorrichtungen zur Implementierung eines derartigen Verfahrens anzugeben.

**[0012]** Diese Aufgaben werden durch ein Verfahren bzw. durch eine Vorrichtung mit den Merkmalen gemäß den Patentansprüchen 1 bzw. 10 gelöst. Vorteilhafte Ausführungsformen und Verwendungen der Erfindung

ergeben sich aus den abhängigen Ansprüchen.

**[0013]** Ein erster wichtiger Gesichtspunkt der Erfindung besteht insbesondere darin, nichtstationäre Schallsignale, bestehend aus Leistungs- und Rauschanteilen, mit einer derart hohen Abtastrate zu erfassen, daß vorbestimmte Signalprofile innerhalb des betrachteten Schallsignals genügend Redundanz für eine Rauschreduzierung enthalten. Phoneme bestehen aus einer Folge von periodischen oder annähernd periodischen Wiederholungen. Auf die Begriffe der periodischen bzw. annähernd periodischen Wiederholungen wird gesondert weiter unten eingegangen. Im folgenden wird einheitlich der Begriff der annähernd periodischen Signalprofile verwendet. Die damit erfaßten Zeitreihen von Schallsignalen liefern Wellenformen, die sich zumindest über bestimmte Signalabschnitte des Schallsignals wiederholen und eine zeitlich begrenzte Anwendung des oben genannten, an sich bekannten Konzepts der nichtlinearen Rauschreduzierung erlauben.

**[0014]** Gemäß einem weiteren wichtigen Gesichtspunkt der Erfindung werden innerhalb eines betrachteten Schallsignals annähernd periodische Signalprofile erfaßt und Korrelationen zwischen den Signalprofilen ermittelt, um korrelierte Signalanteile einem Leistungsanteil und unkorrelierte Signalanteile einem Rauschanteil des Schallsignals zuzuordnen.

**[0015]** Ein weiterer wichtiger Gesichtspunkt der Erfindung besteht in der Idee, zeitliche Korrelationen durch geometrische Korrelationen im Zeitverzögerungs-Einbettungsraum zu ersetzen, die durch Umgebungen in diesem Raum ausgedrückt werden. Punkte in diesen Umgebungen ergeben die Information, die zur nichtlinearen Rauschreduzierung des Punkts notwendig sind, für den die Umgebung konstruiert ist.

**[0016]** Gegenstand der Erfindung ist auch die Bereitstellung einer Vorrichtung zur Signalverarbeitung für Schallsignale, die insbesondere eine Abtastschaltung zur Signalwerterfassung, eine Rechenschaltung zur Signalwertverarbeitung und eine Ausgabeeinheit zur Ausgabe entrauschter Zeitreihen umfaßt.

**[0017]** Schließlich ist hervorzuheben, daß erstmalig die Anwendung nichtlinearer Rauschreduzierungsverfahren für deterministische Systeme zur Verarbeitung nichtstationärer und nicht-deterministischer Schallsignale beschrieben wird. Dies ist überraschend, da die Voraussetzung der an sich bekannten Rauschreduzierungsverfahren insbesondere die Stationarität und den Determinismus der zu bearbeitenden Signale umfassen. Gerade diese Voraussetzungen sind bei nichtstationären Schallsignalen bei Betrachtung des globalen Signalverlaufs verletzt. Dennoch liefert die auf bestimmte Signalprofile beschränkte Anwendung der nichtlinearen Rauschreduzierung hervorragende Ergebnisse.

**[0018]** Die Erfindung besitzt die folgenden Vorteile. Es wird erstmalig ein Rauschreduzierungsverfahren für Schallsignale geschaffen, das im wesentlichen verzerrungsfrei arbeitet und mit einem geringen gerätetechni-

schen Aufwand implementierbar ist. Die Erfindung läßt sich in Echtzeit oder nahezu in Echtzeit implementieren. Bestimmte Teile der erfindungsgemäßen Signalverarbeitung sind mit herkömmlichen Rauschreduzierungsverfahren kompatibel, so daß an sich bekannte zusätzliche Korrekturverfahren oder schnelle Datenverarbeitungsalgorithmen problemlos auf die Erfindung übertragbar sind. Die Erfindung erlaubt die effektive Trennung von Leistungs- und Rauschanteilen unabhängig vom Frequenzspektrum des Rauschens. So ist insbesondere auch sogenanntes farbiges Rauschen oder isospektrales Rauschen trennbar. Die Erfindung ist nicht nur bei stationärem Rauschen, sondern auch beim nichtstationärem Rauschen anwendbar, wenn die Zeitskala, auf der der Rauschprozeß seine Eigenschaften verändert, länger als typischerweise 100 ms ist (dies ist ein Beispielwert, der sich insbesondere auf die Bearbeitung von Sprachsignalen bezieht und bei anderen Anwendungen auch kürzer ausfallen kann).

**[0019]** Die Erfindung ist nicht auf die menschliche Sprache beschränkt, sondern auch bei anderen Schallquellen natürlichen oder synthetischen Ursprungs anwendbar. Bei der Bearbeitung von Sprachsignalen ist es möglich, ein menschliches Sprachsignale von Hintergrundgeräuschen zu trennen. Es ist allerdings nicht möglich, einzelne Sprachsignale voneinander zu trennen. Dies würde voraussetzen, daß z.B. eine Stimme als Leistungsanteil und eine andere Stimme als Rauschanteil betrachtet wird. Die den Rauschanteil repräsentierende Stimme würde jedoch ein nicht behandelbares nichtstationäres Rauschen gleicher Zeitskala darstellen.

**[0020]** Weitere Einzelheiten und Vorteile der Erfindung werden im folgenden unter Bezug auf die beigegeführten Abbildungen beschrieben. Es zeigen:

- Fig. 1 Kurvendarstellungen zur Illustration eines Sprachsignals;
- Fig. 2 eine Kurvendarstellung eines Zeitausschnitts des in Fig. 1 illustrierten Schallsignals;
- Fig. 3 ein Flußdiagramm zur Illustration des erfindungsgemäßen Verfahrens;
- Fig. 4 Kurvendarstellungen zur Illustration einer erfindungsgemäßen Rauschreduzierung an einem Pfeifsignal;
- Fig. 5 Kurvendarstellungen zur Illustration des erfindungsgemäßen Verfahrens an Sprachschallsignalen;
- Fig. 6 eine Darstellung der Rauschreduzierung in Abhängigkeit vom Rauschpegel;
- Fig. 7 eine Kurvendarstellung zur Illustration von

Korrelationen zwischen Signalprofilen in einem Sprachsignal;

- Fig. 8 eine Kurvendarstellung zur Illustration eines rauschbereinigten Sprachsignals;
- Fig. 9 eine schematische Blockdarstellung einer erfindungsgemäßen Vorrichtung; und
- Fig. 10 Kurvendarstellungen zur Illustration der nichtlinearen Rauschreduzierung in deterministischen Systemen (Stand der Technik).

**[0021]** Die Erfindung wird im folgenden am Beispiel der Rauschreduzierung an Sprachsignalen durch Ausnutzung der intra-phonem-Redundanz erläutert. Der Leistungsanteil des Schallsignals wird durch einen Sprachanteil  $x$  gebildet, der durch einen Rauschanteil  $r$  überlagert ist. Das Schallsignal ist in Signalabschnitte unterteilt, die beim Sprachbeispiel durch gesprochene Silben oder Phoneme gebildet werden. Die Erfindung ist jedoch nicht auf die Sprachverarbeitung beschränkt. Bei anderen Schallsignalen wird die Zuordnung der Signalabschnitte anwendungsabhängig anders gewählt. Der erfindungsgemäßen Signalverarbeitung ist jedes Schallsignal zugänglich, das an sich zwar nichtstationär ist, aber innerhalb vorbestimmter Signalabschnitte annähernd periodisch sich wiederholende Signalprofile aufweist.

#### Nichtlineare Rauschreduzierung in deterministischen Systemen

**[0022]** Im folgenden werden zunächst Einzelheiten der nichtlinearen Rauschreduzierung erläutert, wie sie an sich aus den oben zitierten Publikationen von E. J. Kostelich et al. und P. Grassberger et al. bekannt sind. Diese Erläuterungen dienen dem Verständnis der herkömmlichen Technik. In Bezug auf Einzelheiten der nichtlinearen Rauschreduzierung werden hier die genannten Publikationen von E. J. Kostelich et al. und P. Grassberger et al. vollständig in die vorliegende Beschreibung einbezogen. Die Erläuterung bezieht sich auf deterministische Systeme. Die erfindungsgemäße Übertragung der herkömmlichen Technik auf nichtdeterministische Systeme wird unten beschrieben.

**[0023]** Die Zustände  $x$  eines dynamischen Systems werden durch eine Bewegungsgleichung gemäß  $x_{n+1} = F(x_n)$  in einem Zustandsraum beschrieben. Die Bewegungsgleichung ist in der Regel eine komplizierte Differentialgleichung. Falls die Funktion  $F$  nicht bekannt ist, kann sie jedoch aus langen Zeitreihen  $\{x_k\}$ ,  $k = 1, \dots, N$ , durch Betrachtung sämtlicher Punkte in einer Umgebung (oder: Nachbarschaft)  $U_n$  eines Punktes  $x_n$  und Minimierung der Funktion (1) linear angenähert werden.

$$s_n^2 = \sum_{k: x_k \in U_n} (A_n x_k + b_n - x_{k+1})^2, \quad (1)$$

**[0024]** Die Größe  $s_n^2$  stellt einen Vorhersagefehler in Bezug auf die Faktoren  $A_n$  und  $b_n$  dar. Der implizite Ausdruck  $A_n x_k + b_n - x_{k+1} = 0$  illustriert, daß die Werte, die der oben genannten Bewegungsgleichung entsprechen, auf eine Hyperebene innerhalb des betrachteten Zustandsraumes beschränkt sind.

**[0025]** Wenn der Zustand  $x_k$  durch ein statistisches Rauschen  $r_k$  zu einem realen Zustand  $y_k = x_k + r_k$  überlagert ist, so sind die zur Umgebung  $U_n$  gehörenden Punkte nicht mehr auf die durch  $A_n$  und  $b_n$  gebildete Hyperebene beschränkt, sondern in einem Bereich um die Hyperebene gestreut. Die nichtlineare Rauschreduzierung bedeutet nun, die verrauschten Vektoren  $y_n$  auf diese Hyperebene zu projizieren. Die Projektion der Vektoren auf die Hyperebene wird mit an sich bekannten Verfahren der linearen Algebra durchgeführt.

**[0026]** Bei Zeitreihen, wie bei Sprachsignalen, wird nur eine Folge von Skalarwerten aufgezeichnet. Aus diesen werden die zu rekonstruierenden Phasenraumvektoren mit dem Konzept der Zeitversatzvektoren ermittelt, wie es im einzelnen von F. Takens unter dem Titel "Detecting Strange Attractors in Turbulence" in "Lecture Notes in Math", Bd. 898, Springer, New York, 1981, oder von T. Sauer et al. in "J. Stat. Phys.", Bd. 65, 1991, S. 579, und illustrativ im folgenden beschrieben wird. Auch diese Publikationen werden hiermit vollständig in die vorliegende Beschreibung einbezogen.

**[0027]** Ausgehend von einer skalaren Zeitreihe  $s_k$  werden Zeitversatzvektoren in einem m-dimensionalen Raum gemäß  $\hat{s}_n = (s_n, s_{n-\tau}, \dots, s_{n-(m-1)\tau})$  gebildet. Der Parameter m ist die Einbettungsdimension der Zeitversatzvektoren. Die Einbettungsdimension wird anwendungsabhängig gewählt und ist größer als der doppelte Wert der fraktalen Dimension des Attraktors des betrachteten dynamischen Systems. Der Parameter  $\tau$  ist ein Abtastabstand (oder: "time lag"), der den zeitlichen Abstand der aufeinanderfolgenden Elemente der Zeitreihe darstellt. Der Zeitversatzvektor ist somit ein m-dimensionaler Vektor, dessen Komponenten einen bestimmten Zeitreihenwert und die (m-1) vorhergehenden Zeitreihenwerte umfassen. Er beschreibt die zeitliche Entwicklung des Systems während eines Zeitbereichs oder Einbettungsfensters der Dauer  $m \cdot \tau$ . Mit jedem neuen Abtastwert verschiebt sich das Einbettungsfenster innerhalb der gesamten zeitlichen Entwicklung jeweils um einen Abtastabstand. Der Abtastabstand  $\tau$  ist wiederum eine anwendungsabhängig gewählte Größe. Falls sich das System wenig ändert, kann der Abtastabstand größer gewählt werden, um die Verarbeitung redundanter Daten zu vermeiden. Falls sich das System schnell ändert, muß der Abtastabstand geringer gewählt werden, da andernfalls die

Korrelationen, die zwischen benachbarten Werten auftreten, Fehler in die weitere Verarbeitung einführen würden. Die Wahl des Abtastabstandes  $\tau$  ist somit ein Kompromiß zwischen der Redundanz und der Korrelation zwischen aufeinanderfolgenden Zuständen.

**[0028]** Die oben genannte Projektion der Zustände auf die Hyperebene erfolgt unter Verwendung der Zeitversatzvektoren entsprechend einer Berechnung, die im einzelnen von H. Kantz et al. in "Phys. Rev. E", Bd. 48, 1993, S. 1529, beschrieben ist. Auch diese Publikation wird vollständig in die vorliegende Beschreibung einbezogen. Für jeden Zeitversatzvektor  $\hat{s}_n$  werden alle Nachbarn im Zeitverzögerungsraum betrachtet, d.h. die Umgebung  $U_n$  wird gebildet. Anschließend wird die Kovarianzmatrix gemäß Gleichung (2) berechnet, wobei das Zeichen  $\wedge$  bedeutet, daß der Mittelwert auf der Umgebung  $U_n$  subtrahiert worden ist.

$$C_{ij} = \sum_{U_n} (\hat{s}_k)_i (\hat{s}_k)_j \quad (2)$$

**[0029]** Zur Kovarianzmatrix  $C_{ij}$  werden die Singulär- oder Eigenwerte ermittelt. Die Vektoren, die den größten Singulärwerten entsprechen, stellen die Richtungen dar, die die Hyperebene, die durch die oben genannten  $A_n$  und  $b_n$  definiert ist, aufspannen.

**[0030]** Um nun das Rauschen aus den Werten  $\hat{s}_n$  zu reduzieren, werden die zugehörigen Zeitversatzvektoren auf die dominanten Richtungen, die die Hyperebene aufspannen, projiziert. Für jedes Element der skalaren Zeitreihe ergibt dies m verschiedene Korrekturen, die in geeigneter Weise kombiniert werden. Mit den rauschreduzierten Werten kann der beschriebene Vorgang zur erneuten Projektion wiederholt werden.

**[0031]** Die Identifizierung der Nachbarn, die Berechnung der Kovarianzmatrix und Ermittlung von dominanten Vektoren, die einer vorbestimmten Anzahl Q von größten Singulärwerten entsprechen, repräsentiert die Suche nach Korrelationen zwischen aufeinanderfolgenden Systemzuständen. Diese Suche wird bei den deterministischen Systemen auf die bekannte oder angenommene Bewegungsgleichung des Systems bezogen. Wie die erfindungsgemäße Suche nach Korrelationen zwischen Systemzuständen bei nicht-deterministischen Systemen erfolgt, wird unten beschrieben.

#### Nicht-lineare Rauschreduzierung in nicht-deterministischen Systemen

**[0032]** Um die Korrelation zwischen den Zuständen zu ermitteln, wird beim deterministischen System die angenommene zeitliche Unveränderlichkeit der Bewegungsgleichung als Zusatzinformation herangezogen. Im Unterschied dazu erfolgt die Ermittlung der Korrelation zwischen benachbarten Zuständen bei der erfindungsgemäßen Signalverarbeitung im nicht-deter-

ministischen System auf der Grundlage der folgenden Zusatzinformation.

**[0033]** Die Erfindung basiert auf der Ausnutzung von Redundanz im Signal. Wegen der Nicht-Stationarität ist zwischen einer wirklichen Redundanz und zufälligen Ähnlichkeiten von Signalteilen, die jedoch unkorreliert sind, zu unterscheiden. Dies wird durch die Verwendung einer höheren Einbettungsdimension und eines größeren Einbettungsfensters erzielt, als es notwendig wäre, um die jeweils aktuellen Dynamiken aufzulösen. Ein Sprachsignal ist eine Verkettung von Phonemen. Jedes einzelne Phonem ist durch eine charakteristische Wellenform gekennzeichnet, die sich mehrfach nahezu unverändert wiederholt. Ein Zeitversatz-Einbettungsvektor, der eine derartige Welle vollständig abdeckt, kann somit eindeutig einem gegebenen Phonem zugeordnet werden, ohne daß eine Mißinterpretation zu einem anderen Phonem mit einer anderen charakteristischen Wellenform auftritt. Innerhalb eines Phonems ändern sich diese Wellenformen in einer bestimmten Weise, so daß keine absolut exakten Wiederholungen auftreten. Wegen der letztgenannten Eigenschaft wird von nahezu periodischen Wiederholungen gesprochen.

**[0034]** Die menschliche Sprache ist eine Aneinanderreihung von Phonemen oder Silben, die in Bezug auf die auftretenden Amplituden und Frequenzen charakteristische Muster besitzen. Diese Muster können beispielsweise durch Beobachtung elektrischer Signale eines Schallwandlers (z.B. Mikrofon) erfaßt werden. Auf mittleren Zeitskalen (z.B. im Rahmen eines Wortes) ist die Sprache nicht stationär und auf langen Zeitskalen (z.B. im Rahmen eines Satzes) hochkomplex, wobei viele aktive Freiheitsgrade und ggf. langreichweitige Korrelationen auftreten. Auf kurzen Zeitskalen (Zeitbereiche, die im wesentlichen der Länge eines Phonems oder einer Silbe entsprechen), treten im Signalverlauf repetetive Muster oder sich wiederholende Signalprofile auf, die im folgenden erläutert werden. Einzelheiten der konkreten Berechnungen werden analog zur herkömmlichen Rauschreduzierung implementiert und können den o.a. Publikationen entnommen werden.

**[0035]** Fig. 1 zeigt als Beispiel den italienischen Gruß "Buon giorno" als Wellenzug. Dabei handelt es sich um die mit einer Abtastfrequenz von 10 kHz aufgezeichnete Signalamplitude mit den (willkürlich normierten) Zeitreihenwerten  $y_n$  in Abhängigkeit von der dimensionslosen Zeitzählskala. Diese Signalamplitude wurde von einer extrem rauscharmen, digitalen Sprachaufzeichnung abgeleitet. Der Gesamtzeitverlauf von  $n=0$  bis  $n=20000$  entspricht einem Zeitbereich von rd. 2 Sekunden.

**[0036]** Bei Darstellung eines Zeitausschnitts des in Fig. 1 gezeigten Amplitudenverlaufs mit extrem gestreckter Zeitskala ergibt sich das Bild in Fig. 2. Es zeigt sich, daß der Amplitudenverlauf innerhalb bestimmter Signalabschnitte (z.B. Phoneme) die illustrierten periodischen Wiederholungen aufweist. Ein

Signalprofil wiederholt sich bei dem dargestellten Beispiel in Zeitintervallen einer Breite von rund 7 ms. Ein besonderer Vorteil der Erfindung besteht darin, daß die Wirksamkeit der erfindungsgemäßen Rauschreduzierung nicht von der absoluten Exaktheit der dargestellten Periodizität abhängt. Es ist möglich, daß keine exakten Wiederholungen auftreten, sondern eine systematische Modifikation der typischen Wellenform eines Signalprofils innerhalb eines Phonems erfolgt. Diese Variation wird jedoch bei dem unten im einzelnen erläuterten Verfahren berücksichtigt, da sie die Freiheit in den nach der Projektion  $Q$  verbleibenden Richtungen repräsentiert. Zur Berücksichtigung der Variation (Abweichung von exakten Wiederholungen) wird hier der Begriff der annähernd periodischen Signalprofile verwendet, die sich von exakt periodischen Signalprofilen lediglich durch eine systematische Variabilität unterscheiden.

**[0037]** Bei einer Zeitversatzeinbettung (mit geeignet gewählten Parametern  $m$  und  $\tau$ , s.o.) bilden die gezeigten Wiederholungen benachbarte Punkte im Zustandsraum (bzw. Vektoren, die auf diese Punkte gerichtet sind). Ist nun die Variabilität in diesen Punkten durch eine Überlagerung durch Rauschen größer als die natürliche Variabilität aufgrund der Nichtstationarität, so wird eine approximative Identifikation der Mannigfaltigkeit und die Projektion darauf das Rauschen stärker reduzieren als es das eigentliche Signal beeinflusst. Dies ist der Grundansatz des erfindungsgemäßen Verfahrens, das im folgenden unter Bezug auf das Flußdiagramm gemäß Fig. 3 erläutert wird.

**[0038]** Fig. 3 ist eine Übersichtsdarstellung, die schematisch grundsätzliche Schritte des erfindungsgemäßen Verfahrens zeigt. Die Erfindung ist jedoch nicht auf diesen Ablauf beschränkt. Anwendungsabhängig kann eine Modifizierung in Bezug auf die Datenaufnahme, die Parameterermittlung, die eigentliche Berechnung zur Rauschreduzierung, die Trennung von Leistungs- und Rauschanteilen und die Ausgabe des Ergebnisses vorgesehen sein.

**[0039]** Gemäß Fig. 3 erfolgt nach dem Start 100 die Datenaufnahme 101 und die Parameterermittlung 102. Die Datenaufnahme 101 umfaßt die Aufnahme eines Schallsignals durch Umwandlung des Schalls in eine elektrische Größe. Die Datenaufnahme kann zur analogen oder digitalen Schallaufzeichnung eingerichtet sein. Anwendungsabhängig wird das Schallsignal in einem Datenspeicher oder bei Echtzeitverarbeitung in einem Pufferspeicher (s. Fig. 9) gespeichert. Die Parameterermittlung 102 umfaßt die Auswahl von Parametern, die für die spätere Suche nach Korrelationen zwischen benachbarten Zuständen im Schallsignal geeignet sind. Diese Parameter umfassen insbesondere die Einbettungsdimension  $m$ , den Abtastabstand  $\tau$ , die  $\epsilon$ -Durchmesser der Umgebungen  $U$  im Zeitversatz-Einbettungsraum zur Identifizierung von Nachbarn, und die Zahl  $Q$  der Zeitversatzvektoren, auf die die Zustandsprojektion erfolgen soll.

**[0040]** Bei der Sprachsignalverarbeitung kann die

Einbettungsdimension  $m$  beispielsweise im Bereich von 10-50, vorzugsweise 20-30, und der Abtastabstand  $\tau$  im Bereich von 0.1 bis 0.3 ms liegen, so daß das Einbettungsfenster  $m \cdot \tau$  vorzugsweise rd. 3 bis 8 ms abdeckt. Diese Daten beziehen sich auf eine Phonemdauer von rd. 50 bis 200 ms und die Komplexität der menschlichen Stimme. Typische Signalprofile liegen wegen der Tonlage der menschlichen Stimme von ungefähr 100 Hz zwischen 3 und 15 ms. Fig. 2 zeigt beispielsweise Wiederholungen des Signalprofils jeweils nach 7 ms. Die Parameterermittlung 102 (Fig. 3) kann im Zusammenspiel mit der Datenaufnahme 101 erfolgen oder im Rahmen einer Vorabanalyse durchgeführt worden sein. Bei einer Vorabanalyse wird die Einbettungsdimension  $m$  und die Dimension der Mannigfaltigkeit (entsprechend dem Parameter  $Q$ ), in der die unverrauschten Daten lägen, abgeschätzt. Es kann auch vorgesehen sein, daß die Parameterermittlung 102 während des Verfahrens wiederholt wird. Dies kann beispielsweise als Korrektur in Reaktion auf das Ergebnis der Leistungs-/Rausch-Trennung 109 (s.u.) erfolgen.

**[0041]** Auf der Grundlage der aufgenommenen Meßwerte und der festgelegten Parameter folgt die Signalabtastung 103. Die Signalabtastung 103 ist dazu vorgesehen, die Werte der Zeitreihen  $y_n$  entsprechend den vorher festgelegten Abtastparametern aus den Daten zu ermitteln. Die folgenden Schritte 104 bis 109 stellen die eigentliche Berechnung der Projektionen der realen Schallsignale auf unverrauschte Schallsignale oder -zustände dar.

**[0042]** Der Schritt 104 umfaßt die Bildung des ersten Zeitversatzvektors am Anfang der Zeitreihen (z.B. gemäß Fig. 2). Dieser erste Zeitversatzvektor muß sich nicht notwendigerweise auf das zeitlich zuerst erscheinende erste Signalprofil beziehen. Dies wird jedoch insbesondere bei Echtzeit- oder Quasiechtzeitverarbeitungen bevorzugt. Der erste Zeitversatzvektor umfaßt  $m$  Signalwerte  $y_n$  als  $m$  Komponenten, die mit dem Zeitversatz  $\tau$  aufeinanderfolgen. Anschließend werden bei Schritt 105 benachbarte Zeitversatzvektoren (Nachbarvektoren) gebildet und erfaßt. Die Nachbarvektoren beziehen sich auf sehr ähnliche Signalprofile wie das Signalprofil, das durch den ersten Vektor repräsentiert wird. Sie bilden die erste Umgebung  $U$ . Wenn der erste Vektor ein Profil repräsentiert, der Teil eines Phonems ist, so entsprechen die Nachbarvektoren im wesentlichen den sich annähernd wiederholenden Signalprofilen innerhalb desselben Phonems. Bei der Sprachverarbeitung wiederholen sich etwa 15 Signalprofile innerhalb eines Phonems. Die Zahl der ermittelten Nachbarvektoren ist kleiner oder gleich der Zahl sich wiederholender Signalprofile und beträgt beispielsweise rund 5 bis 15.

**[0043]** Anschließend folgt die Berechnung der Kovarianzmatrix 106 entsprechend der oben angegebenen Gleichung (2). Die in diese Matrix eingefügten Vektoren sind die Vektoren aus der Basis-Umgebung  $U$  entsprechend der Festlegung in Schritt 105. Schritt 106

umfaßt dann die Bestimmung der  $Q$  größten Singulärwerte der Kovarianzmatrix und der zugehörigen Singulärvektoren im  $m$ -dimensionalen Raum.

**[0044]** Im Rahmen der folgenden Projektion 107 werden alle Anteile des ersten Zeitversatzvektors, die nicht in dem von den  $Q$  ermittelten dominanten Vektoren aufgespannten Unterraum liegen, eliminiert. Der Wert  $Q$  liegt im Bereich von rund 2 bis 10, vorzugsweise 4 bis 6. Bei einem modifizierten Verfahren kann der Wert  $Q$  Null betragen (s. unten).

**[0045]** Die verhältnismäßig geringe Zahl  $Q$ , die die Dimension des Unterraumes repräsentiert, auf die die Zustände bzw. Signale projiziert werden, stellt einen besonderen Vorteil der Erfindung dar. Es wurde festgestellt, daß der Dynamikbereich der Wellen innerhalb eines gegebenen Phonems nur relativ wenige Freiheitsgrade besitzt, wenn sie einmal innerhalb eines hochdimensionalen Raumes identifiziert worden ist. Daher sind auch verhältnismäßig wenige Nachbarzustände für die Projektionsberechnung erforderlich. Für die Erfassung der Korrelation zwischen den Signalprofilen sind nur die größten Singulärwerte und entsprechenden Singulärvektoren der Kovarianzmatrix relevant. Dieses Ergebnis ist überraschend, da die nichtlineare Rauschreduzierung an sich für deterministische Systeme mit umfangreichen Zeitreihen entwickelt wurde. Außerdem ergibt sich als besonderer Vorteil ein verhältnismäßig geringer Zeitaufwand für die Berechnung.

**[0046]** Anschließend wird der nächste Zeitversatzvektor bei Schritt 108 ausgewählt und die Schrittfolge 105-107 wiederholt, wobei neue Umgebungen und neue Kovarianzmatrizen gebildet werden. Diese Wiederholung erfolgt, bis alle Zeitversatzvektoren, die aus der Zeitreihe konstruiert werden können, verarbeitet worden sind.

**[0047]** Die Bildung bzw. Erfassung der Nachbarvektoren (Schritt 105) erfolgt übrigens bei einer höheren Dimension als die Projektion 107. Die hohe Dimension bei der Nachbarsuche garantiert die Auswahl der richtigen Nachbarn, die Profile repräsentieren, die von denselben Phonemen abgeleitet sind. Die Erfindung wählt somit implizit ohne irgendein Sprachmodell Phoneme aus. Wie oben erklärt wurde, repräsentiert die Dynamik innerhalb eines Phonems jedoch wesentlich weniger Freiheitsgrade, so daß innerhalb des durch die Singulärvektoren aufgespannten Unterraumes niedrigdimensional und schnell gearbeitet werden kann. Für Echtzeitanwendungen erfolgt die Schallsignalverarbeitung im wesentlichen für die Phoneme aufeinanderfolgend, so daß Phonem für Phonem komplett abgearbeitet und so ein entraushtes Ausgangssignal erzeugt wird. Dieses Ausgangssignal ist gegenüber dem erfaßten Schallsignal (Eingangssignal) um rund 100-200 ms verzögert (Echtzeit- oder auch Quasi-Echtzeitanwendung).

**[0048]** Die Schritte 109 und 110 betreffen die Bildung des eigentlichen Ausgangssignals. Der Schritt 109 ist auf die Trennung von Leistungs- und Rauschsi-

gnalen gerichtet. Ein entauschtes Zeitreihenelement  $s_k$  wird durch Mittlung über die korrespondierenden Elemente aus allen Zeitversatzvektoren, die dieses Element enthalten, gebildet. Statt einer einfachen Mittlung kann eine gewichtete Mittlung eingeführt werden. Nach Schritt 109 kann ein Rücksprung vor Schritt 104 vorgesehen sein. Die entauschten Zeitreihenelemente bilden dann die Eingangsgrößen für die erneute Bildung von Zeitversatzvektoren und deren Projektion auf den Unterraum entsprechend den Singulärvektoren. Diese Verfahrenswiederholung ist nicht notwendig, kann aber zur Verbesserung der Rauschreduzierung z.B. 2- oder 3-fach vorgesehen sein. Nach Schritt 109 kann aber auch ein Rücksprung zur Parameterermittlung 102 vorgesehen sein, falls der nach Schritt 109 vorliegende Leistungsanteil sich weniger als erwartet (z.B. durch weniger als einen vorbestimmten Schwellwert) vom unverarbeiteten Schallsignale unterscheidet. Hierzu können im Verfahrensablauf nicht dargestellte Entscheidungsmechanismen eingebaut sein. Bei Schritt 110 folgt die Datenausgabe. Bei der Rauschreduzierung wird als Leistungsanteil das rauschreduzierte Sprachsignal ausgegeben. Alternativ kann aber auch anwendungsabhängig gerade die Ausgabe oder Speicherung des Rauschanteils vorgesehen sein.

**[0049]** Die oben erläuterte Verfahrensweise kann in Bezug auf die Parameterbestimmung unter Berücksichtigung der folgenden Gesichtspunkte modifiziert werden. Erstens kann die Dimension der Mannigfaltigkeit (entsprechend dem Parameter  $Q$ ), in der die rauschfreien Daten liegen würden, im Verlauf eines Signals variieren. Die Dimension  $Q$  kann von Phonem zu Phonem variieren. Die Dimension kann beispielsweise auch während einer Pause zwischen zwei gesprochenen Worten oder irgendeiner anderen Ruhephase Null betragen. Zweitens ist eine Auswahl von relevanten inhärenten Zeitversatzvektoren, auf die der Zustand projiziert werden soll, ausgeschlossen, falls das Rauschen relativ hoch ist (ungefähr 50%). In diesem Fall würden alle Eigenwerte der Korrelationsmatrix ungefähr gleich sein.

**[0050]** Dementsprechend kann beim Verfahrensablauf die folgende Variation des Parameters  $Q$  vorgesehen sein. Anstelle einer unveränderlichen Projektionsdimension  $Q$  wird die Dimension für jede Kovarianzmatrix angepaßt variiert bzw. individuell bestimmt. Bei Schritt 102 wird eine Konstante  $f = 1$  bestimmt. Diese Konstante  $f$  wird empirisch ermittelt. Sie hängt von der Signalart ab und beträgt beispielsweise bei Sprache  $f = 0.1$ . Der maximale Singulärwert einer gegebenen Kovarianzmatrix, multipliziert mit der Konstanten  $f$ , repräsentiert einen Schwellwert. Die Anzahl der Singulärwerte, die größer als der Schwellwert sind, wird dann als Wert für  $Q$  für die Projektion verwendet, vorausgesetzt dieser Wert überschreitet nicht einen bestimmten Maximalwert. Dieser Maximalwert ist z.B. 8. Im letzteren Fall sind alle Singulärwerte einer gegebenen Kovarianzmatrix so ähnlich, daß kein

ausgeprägter linearer Unterraum ausgewählt werden kann und somit  $Q = 0$  gewählt werden muß. Anstelle einer Projektion wird der aktuelle Zeitversatzvektor dann durch den Mittelwert seiner Umgebung ersetzt.

**[0051]** Durch diese Abwandlung wird die Leistungsfähigkeit des Verfahrens insbesondere bei hohen Rauschpegeln drastisch erhöht.

## Beispiele

**[0052]** Im folgenden wird die erfindungsgemäße Signalverarbeitung an zwei Beispielen illustriert. Beim ersten Beispiel ist das verarbeitete Schallsignale ein menschlicher Pfiff (s. Fig. 4). Das zweite Beispiel betrifft die oben genannten Worte "Buon giorno" (s. Fig. 5 bis 8).

**[0053]** Fig. 4 zeigt das Leistungsspektrum für einen menschlichen Pfiff der Dauer 3 s. Ein Pfiff ist ein im wesentlichen periodisches Signal mit charakteristischen Harmonischen und nur geringen Nicht-Stationaritäten. Fig. 4a zeigt den Amplitudenverlauf der Originalaufzeichnung. Nach numerischer Addition eines 10-%igen Rauschens ergibt sich das in Fig. 4b dargestellte Spektrum. Dieses liefert die Eingangsdaten für Schritt 101 des Verfahrensablaufes (Fig. 3). Nach der erfindungsgemäßen Rauschreduzierung ergibt sich das in Fig. 4c gezeigte Bild. Dieses zeigt die vollständige Wiederherstellung des ursprünglichen, unverrauschten Signals. Die Figuren 4a bis 4c zeigen einen besonderen Vorteil der Erfindung gegenüber einem herkömmlichen Filter im Frequenzraum. Ein Filter im Frequenzraum würde alle Leistungsanteile mit Amplituden unterhalb  $10^{-6}$  abschneiden, so daß das entauschte Spektrum nur noch den Peak bei 0 und den Peak um die Grundfrequenz enthielte. Dementsprechend wäre die aus der Rücktransformation gewonnene Zeitreihe völlig ohne Oberschwingungen, was sich sehr synthetisch anhören würde. Diese Nachteile werden bei der erfindungsgemäßen Rauschreduzierung vermieden.

**[0054]** Fig. 5 zeigt entsprechende Ergebnisse am Beispiel von Kurvendarstellungen für die Verarbeitung von Sprachsignalen. In Fig. 5a ist ein Ausschnitt aus dem unverrauschten Wellenzug der Worte "Buon giorno" bezogen auf den Signalverlauf gemäß Fig. 1 analog zu Fig. 2 gezeigt. Es ist die zeitlich begrenzte Wiederholung von Signalprofilen erkennbar, die die zur Reduktion des Rauschens notwendige Redundanz enthält. Fig. 5b zeigt den Wellenzug nach Addition eines synthetischen Rauschens. Nach der erfindungsgemäßen Rauschreduzierung ergibt sich das Bild gemäß Fig. 5c. Es zeigt sich, daß das Originalsignal zum größten Teil rekonstruiert werden konnte.

**[0055]** Die Funktionsfähigkeit der erfindungsgemäßen Rauschreduzierung wurde in verschiedenen Rauscharten und -amplituden geprüft. Als Maß für die Leistungsfähigkeit der Rauschreduzierung kann die Dämpfung  $D$  (in dB) gemäß Gleichung (3) betrachtet werden.

$$D = 10 \log ((\Sigma(\hat{y}_k - x_k)^2) / (\Sigma(y_k - x_k)^2)) \quad (12)$$

**[0056]** In Gleichung (3) steht  $X_k$  für das unverrauschte Signal (Leistungsanteil),  $y_k$  für das ver-  
rauschte Signal (Eingangs-Schallsignal) und  $\hat{y}_k$  für das  
Signal nach der erfindungsgemäßen Rauschreduzierung.

**[0057]** Fig. 6 illustriert die Abhängigkeit der Dämpfung  $D$  der nichtlinear Rauschreduzierung in Abhängigkeit von der relativen Rauschamplitude (Varianz des Rauschanteils : Varianz des Leistungsanteils). Es zeigt sich, daß die Dämpfung selbst bei relativ hohen Rauschamplituden (im Bereich von mehr als 100%) verstärkt ist.

**[0058]** Die Figuren 7 und 8 zeigen weitere Einzelheiten der Sprachrauschreduzierung. Fig. 7 illustriert das Auftreten sich wiederholender Signalprofile innerhalb des im oberen Teil der Figur gezeigten Phonemzuges. In Abhängigkeit von einem (willkürlichen) Zeitindex  $i$  ist im unteren Teil der Figur ein Kurvenbild gedruckt, das aus unter den folgenden Bedingungen gebildeten Punkten besteht. Für jeden Zeitpunkt  $i$  wird der zugehörige Zeitversatzvektor  $\hat{s}_i$  und die Menge aller Zeitversatzvektoren  $\hat{s}_{j,i}$  betrachtet. Falls der Betrag des Differenzvektors zwischen dem  $\hat{s}_i$  und jedem  $\hat{s}_j$  kleiner als eine vorbestimmte Grenze ist, wird ein Punkt gedruckt. Die Punkte bilden mehr oder wenige ausgedehnte Linien. Die Linienstrukturen zeigen, daß innerhalb der Phoneme die oben erläuterten Periodizitäten der Signalprofile auftreten. Die Lücken in diesen Liniensegmenten zeigen, daß die Umgebungen geeignet sind, zwischen verschiedenen Phonehmen zu differenzieren. Für Linienstrukturen, die in Ordinateurichtung besonders ausgedehnt sind, gilt, daß die Anzahl von intra-Phonem-Nachbarn besonders groß ist. Es zeigt sich jedoch auch, daß in der Regel für  $|i-j| > 2000$  keine Wiederholungen auftreten.

**[0059]** Fig. 8 zeigt wiederum am Beispiel der Worte "Buon giorno" im oberen Teil der Figur das unverrauschte Signal, im mittleren Teil das synthetisch hinzugefügte Rauschen und im unteren Teil das nach der Rauschreduzierung verbliebene Rauschen. Die Ordinateurichtung ist in allen drei Fällen identisch. Das verbleibende Rauschen (unterster Teil der Figur) zeigt eine systematische Variation, die darauf hinweist, daß der Erfolg der erfindungsgemäßen Rauschreduzierung selbst vom Schallsignal, d.h. vom konkreten Phonem, abhängt.

**[0060]** Gegenstand der Erfindung ist auch eine Vorrichtung zur Implementierung des erfindungsgemäßen Verfahrens. Gemäß Fig. 9 umfaßt eine Rauschreduzierungsanordnung einen Meßwertaufnehmer 91, einen Datenspeicher 92 und/oder einen Pufferspeicher 93, eine Abtastschaltung 94, eine Rechenschaltung 95 und eine Ausgabereinheit 96.

**[0061]** Die hier vorgestellten Komponenten der erfindungsgemäßen Vorrichtung werden vorzugsweise als festverschaltete Schaltkreisordnung oder als inte-

rierter Chip hergestellt.

**[0062]** Im folgenden werden bevorzugte Anwendungen der Erfindung genannt. Neben der bereits genannten Rauschreduzierung an Sprachsignalen ist die Erfindung auch bei der Rauschreduzierung an Hörgeräten und zur Verbesserung der computergestützten automatischen Spracherkennung anwendbar. In Bezug auf die Spracherkennung kann insbesondere vorgesehen sein, die entrauschten Zeitreihenwerte bzw. Sektoren mit Tabellenwerten zu vergleichen. Die Tabellenwerte repräsentieren entsprechende Werte bzw. Vektoren vorbestimmter Phoneme. Eine automatische Spracherkennung kann somit mit dem Rauschreduzierungsverfahren integriert werden.

**[0063]** Weitere Anwendungen liegen im Bereich der Telekommunikation und bei der Signalverarbeitung anderer Schallquellen als der menschlichen Sprache. Hierzu zählen beispielsweise Tierstimmen oder auch Musik.

## Patentansprüche

1. Verfahren zur Signalverarbeitung eines Schallsignals  $y$ , bei dem innerhalb vorbestimmter Signalabschnitte des Schallsignals redundante Signalprofile erfaßt und Korrelationen zwischen den Signalprofilen ermittelt werden, wobei korrelierte Signalanteile einem Leistungsanteil und unkorrelierte Signalanteile einem Rauschanteil des Schallsignals zugeordnet werden.
2. Verfahren gemäß Anspruch 1, bei dem die Korrelationen zwischen den Signalprofilen mit Verfahren der nichtlinearen Rauschreduzierung in deterministischen Systemen ermittelt werden.
3. Verfahren gemäß Anspruch 1, bei dem das Schallsignal  $y$ , das sich aus einem Sprachanteil  $x$  und einem Rauschanteil  $r$  zusammensetzt in jedem Signalabschnitt gemäß den folgenden Schritten verarbeitet wird:

a) Erfassung einer Vielzahl von Schallsignalwerten  $y_k = x_k + r_k$  mit einem Abtastabstand  $\tau$ ,

b) Bildung von Zeitversatzvektoren, von denen jeder aus Komponenten  $y_k$  besteht, deren Anzahl  $m$  eine Einbettungsdimension und deren Indizes  $k$  vom Einbettungsfenster der Breite  $m \cdot \tau$  abgeleitet sind, wobei für jeden dieser Zeitversatzvektoren eine Umgebung  $U$  aus allen Zeitversatzvektoren gebildet wird, deren Abstand zu einem gegebenen Zeitversatzvektor kleiner als ein vorbestimmter Wert  $\epsilon$  ist;

c) Ermittlung von Korrelationen zwischen den Zeitversatzvektoren und Projektion der Zeitversatzvektoren auf vorbestimmte Singulärwert-

vektoren, und

d) Ermittlung von Nutzsinalwerten, die ein Sprachsignal, das im wesentlichen dem Sprachanteil  $x_k$  entspricht, und/oder ein Rauschanteil  $r_k$  entspricht. 5

4. Verfahren gemäß Anspruch 3, bei der die Zahl  $k$  der gebildeten Zeitversatzvektoren, die die Umgebung bilden, von der Redundanz abhängt, die in den annähernden Wiederholungen der Signalprofile gespeichert ist. 10
5. Verfahren gemäß Anspruch 3, bei dem die Korrelationen zwischen den Zeitversatzvektoren durch die Identifizierung der Umgebung und durch eine Berechnung einer Kovarianzmatrix auf den Vektoren extrahiert werden, die zu der Umgebung gehören. 15
6. Verfahren gemäß Anspruch 3, bei dem die Schritte b) bis c) mit den Nutzsinalwerten mindestens einmal wiederholt werden, wobei zur Verbesserung des Ergebnisses das Verfahren wiederholt werden kann, wenn die gesamte Zeitreihe entrauscht ist. 20 25
7. Verfahren gemäß Anspruch 3, bei dem das Schallsignal ein Sprachsignal ist.
8. Verfahren gemäß einem der Ansprüche 3 bis 7, bei dem das Einbettungsfenster  $m \cdot \tau$  im Bereich von 1 bis 20 ms liegt. 30
9. Verfahren gemäß Anspruch 3, bei dem bei Schritt c) die Zeitversatzvektoren auf eine  $Q$ -dimensionale Mannigfaltigkeit mit adaptiv eingestelltem  $Q$ -Wert projiziert werden. 35
10. Vorrichtung zur Durchführung eines Verfahrens gemäß einem der vorhergehenden Ansprüche, die einen Meßwertaufnehmer (91), einen Datenspeicher (92) und/oder einen Pufferspeicher (93), eine Abtastschaltung (94), eine Rechenschaltung (95) und eine Ausgabeeinheit (96) aufweist. 40 45
11. Verwendung eines Verfahrens gemäß einem der Ansprüche 1 bis 9 zur Rauschreduzierung an Sprachsignalen in der Telekommunikation, in Hörgeräten oder bei der automatischen Spracherkennung. 50
12. Verwendung nicht-linearer Rauschreduzierungsverfahren für deterministische Systeme zur Rauschreduzierung von Sprachsignalen. 55

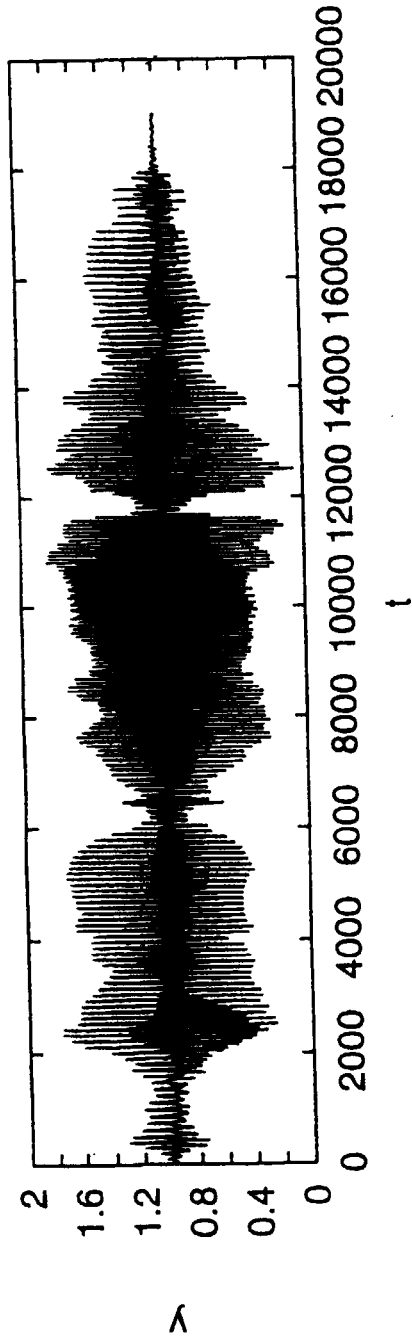


Fig. 1

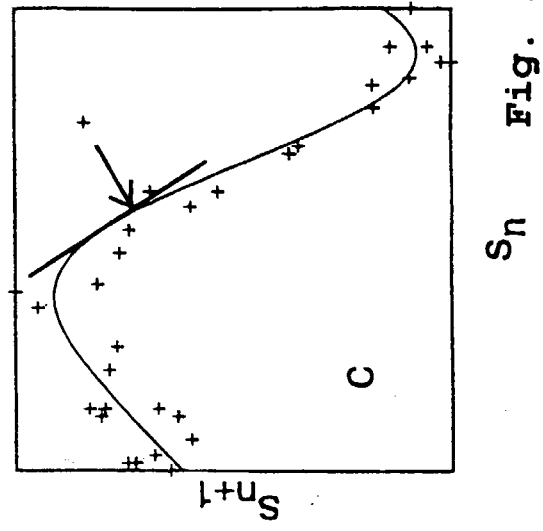
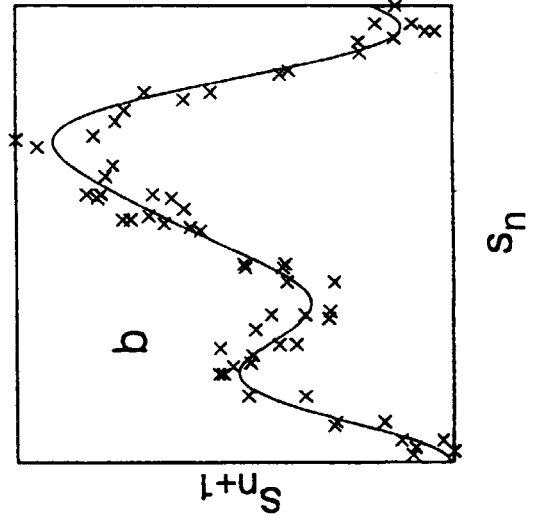
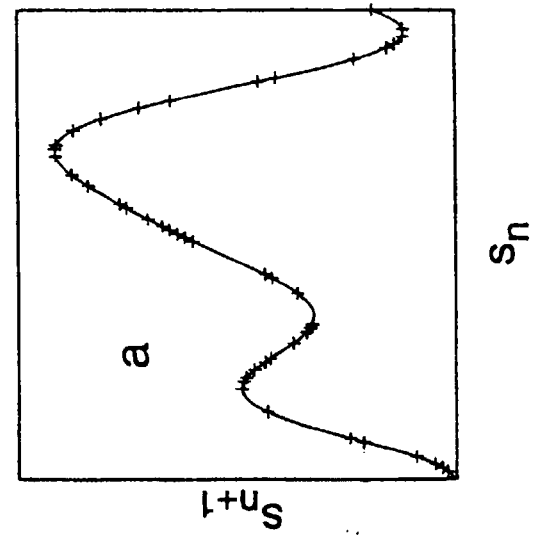


Fig. 10

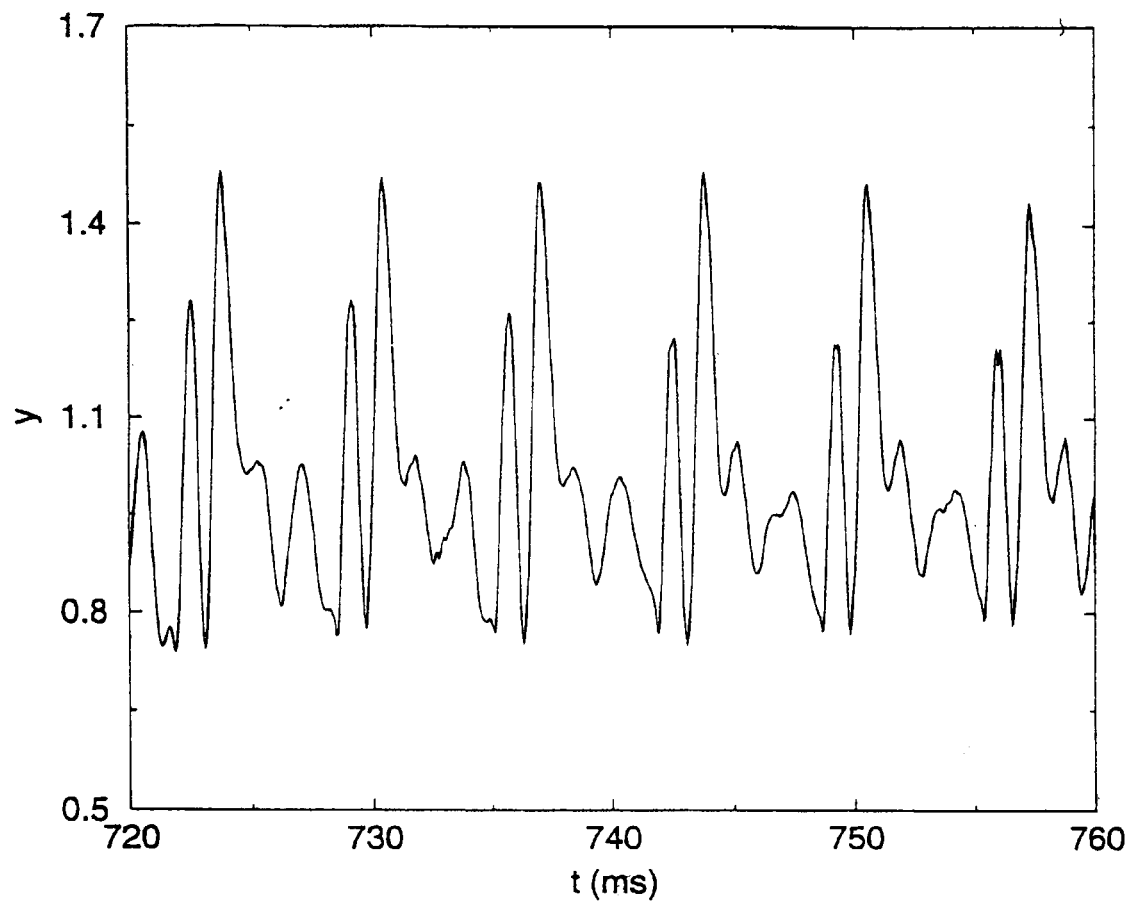


Fig. 2

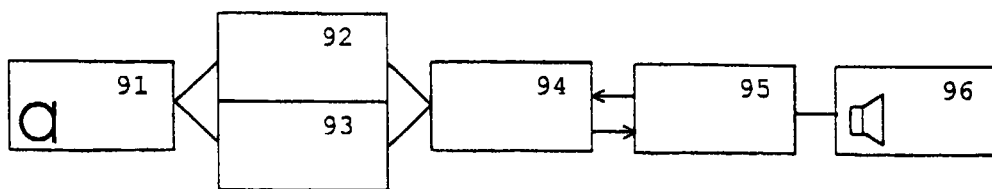


Fig. 9

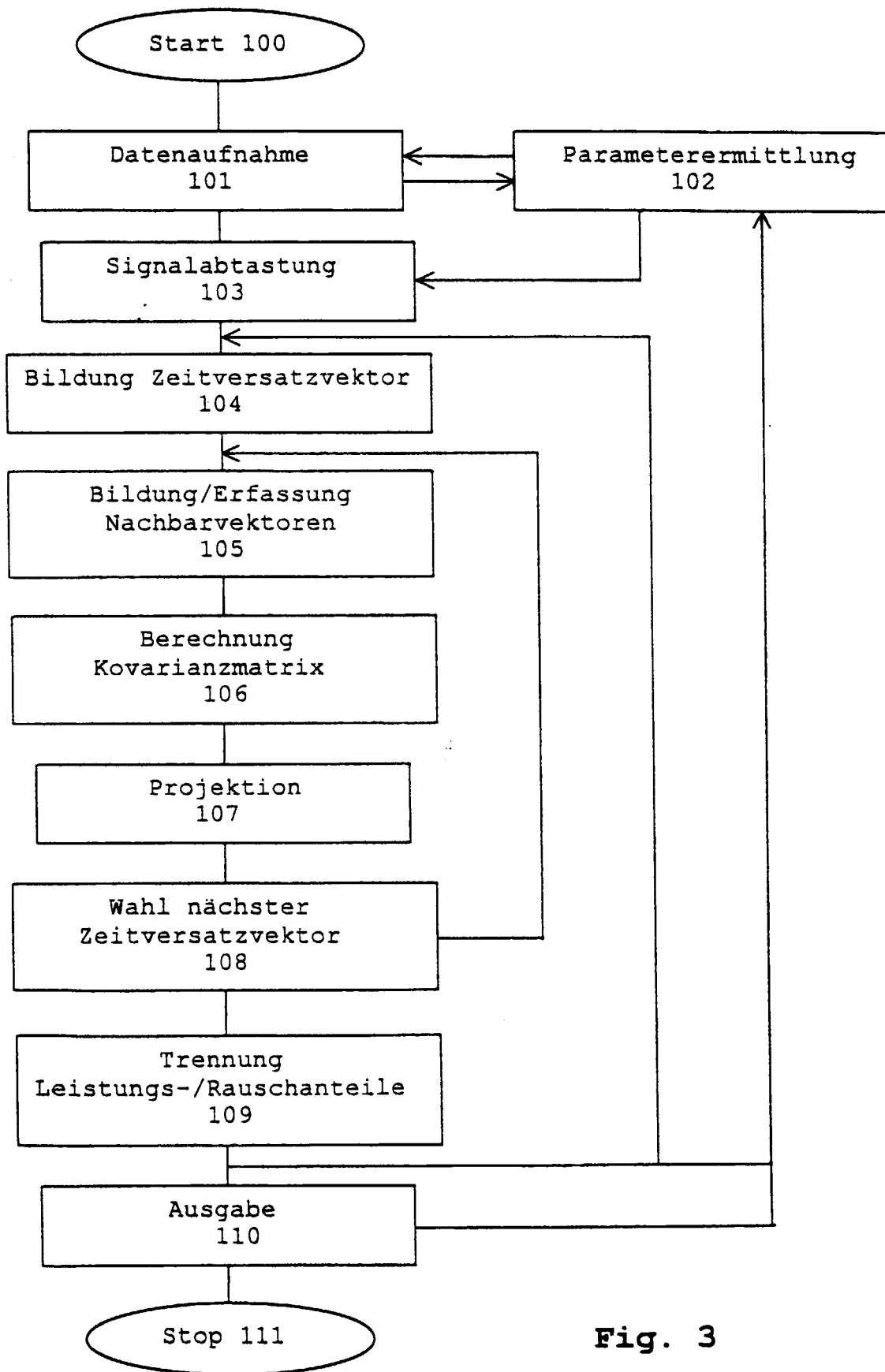


Fig. 3

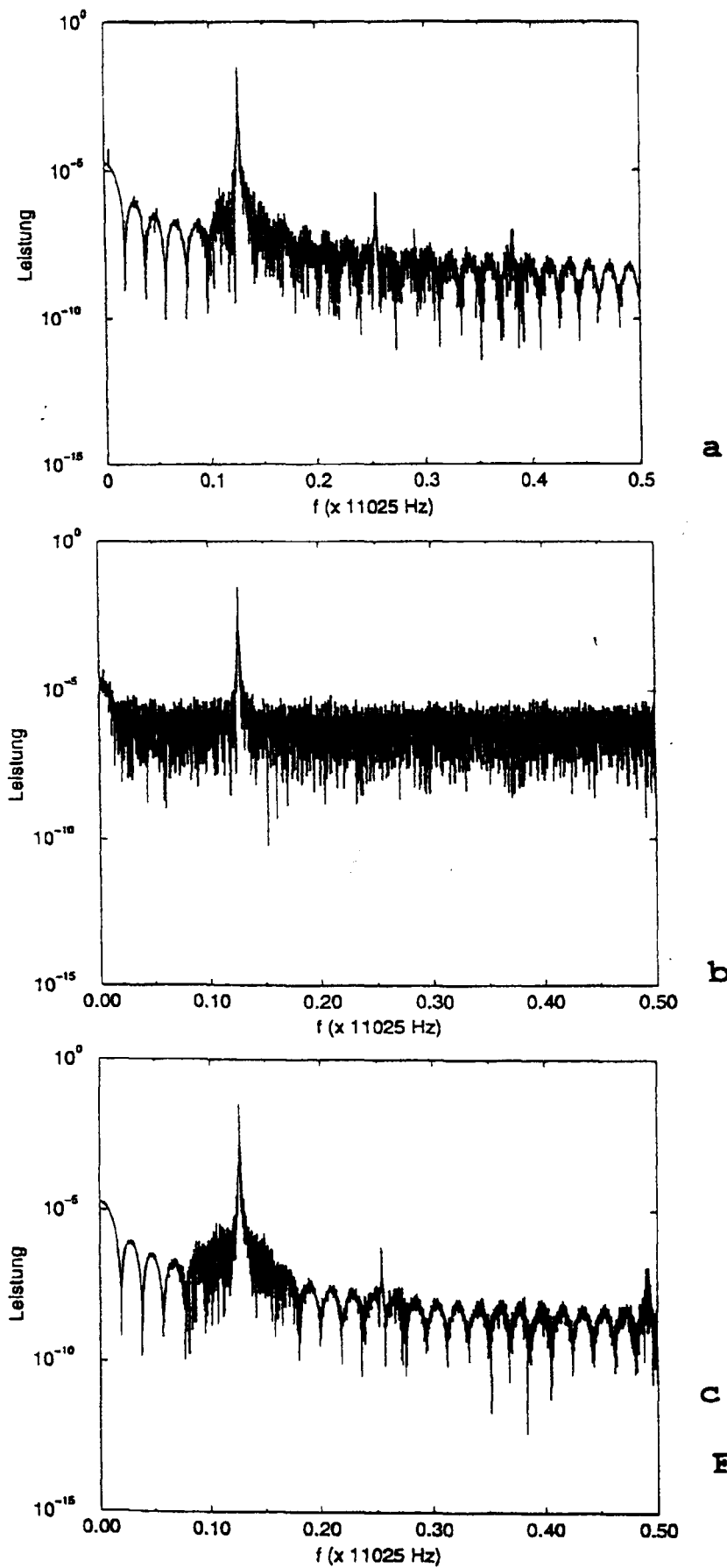


Fig. 4

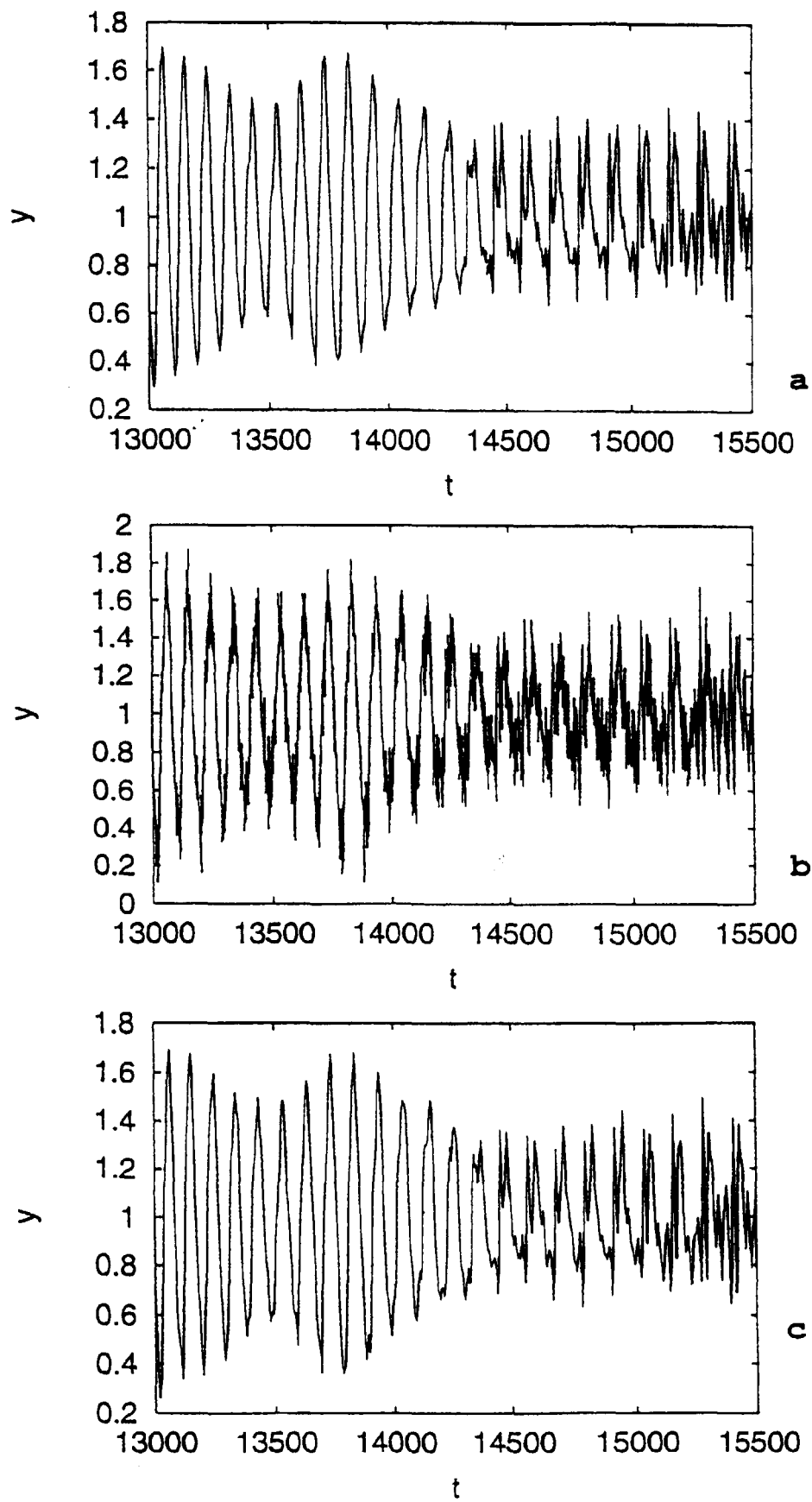


Fig. 5

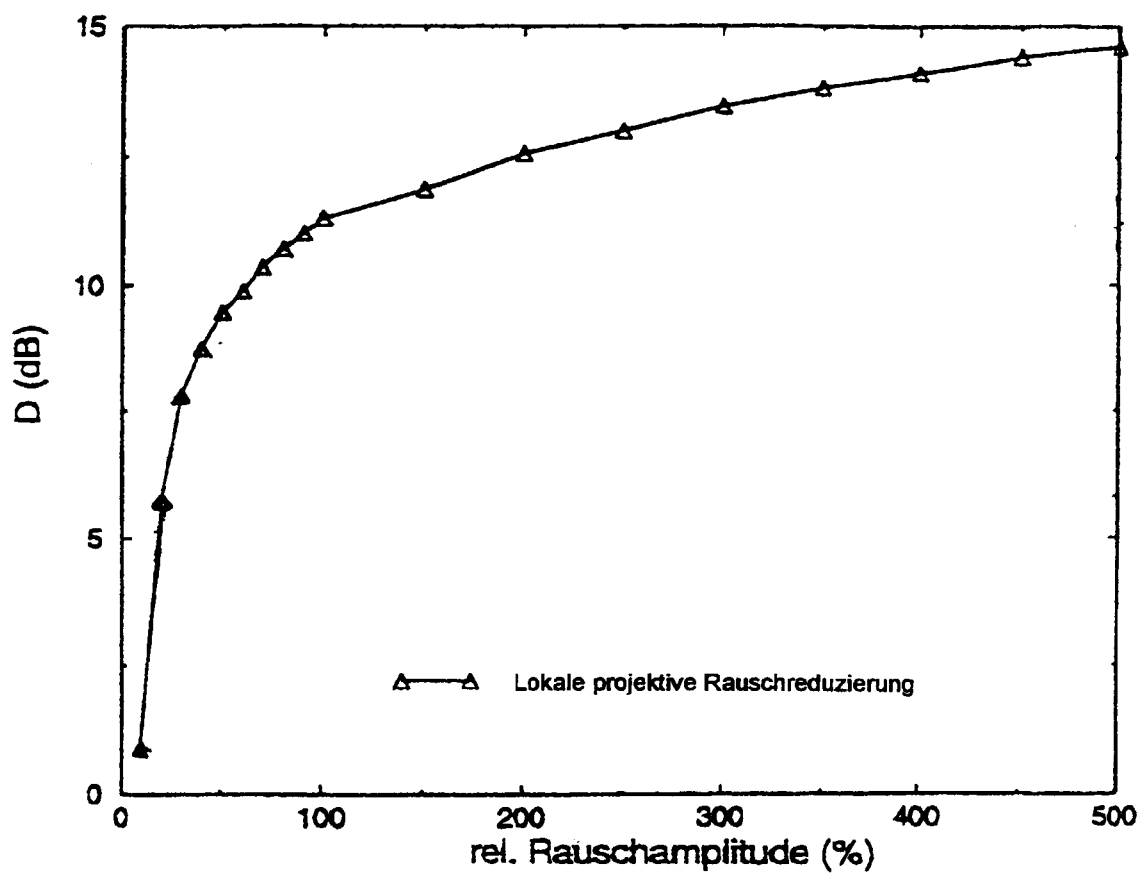


Fig. 6

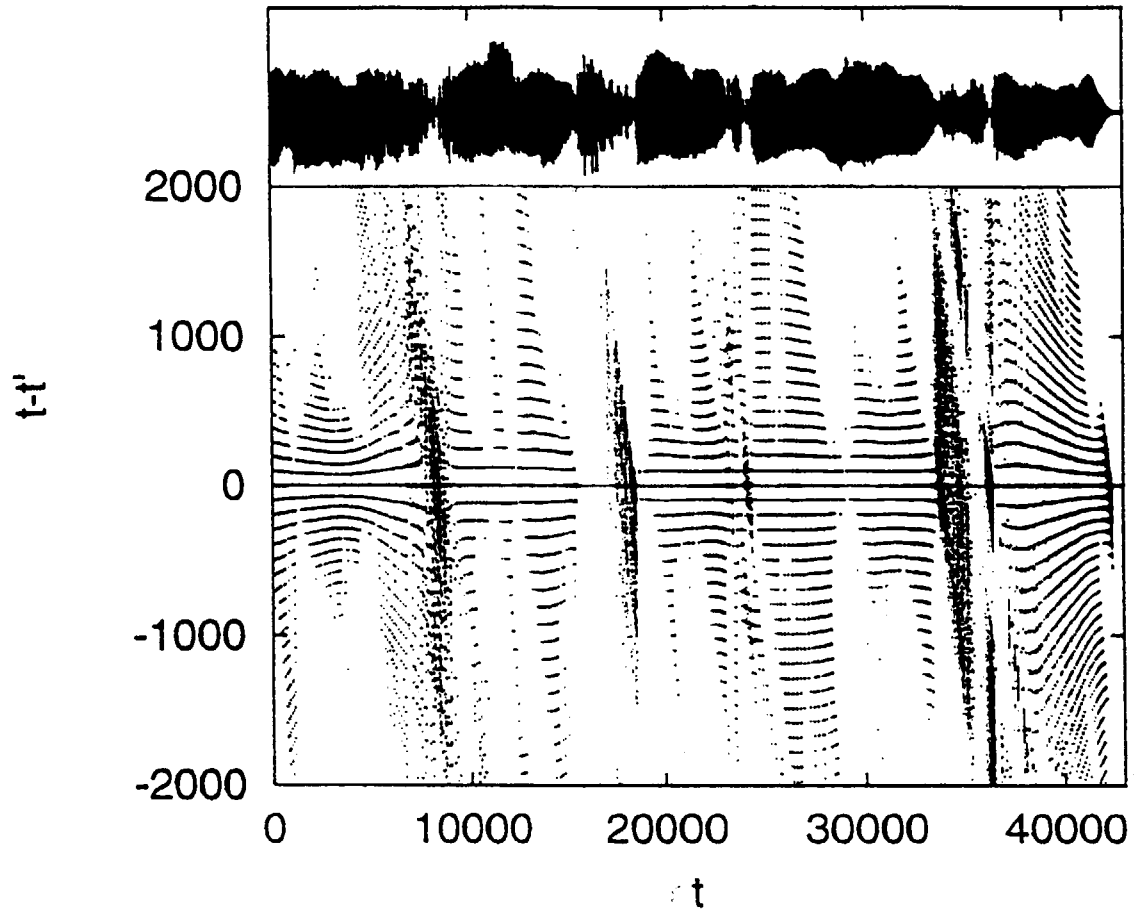
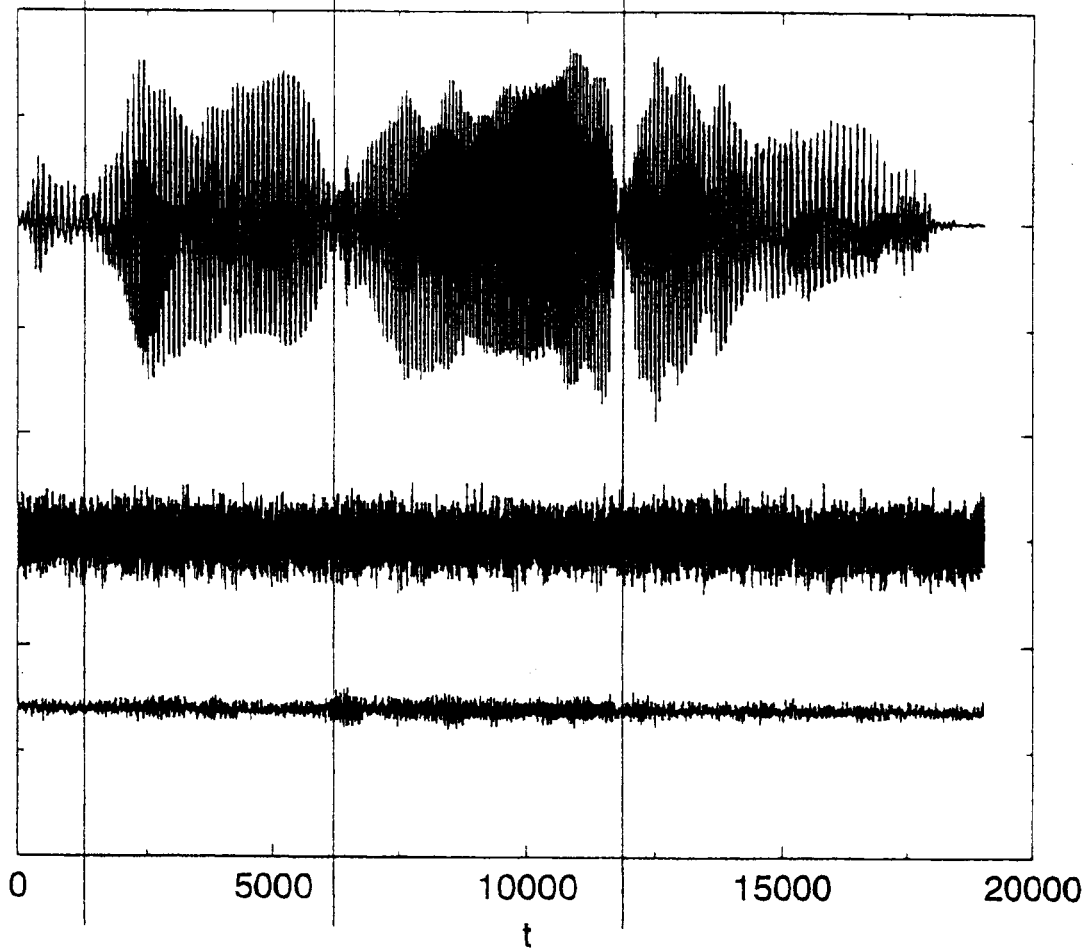


Fig. 7

**bu on gior no**



**Fig. 8**