



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 600 07 633 T2 2004.11.18**

(12)

Übersetzung der europäischen Patentschrift

(97) **EP 1 192 567 B1**

(51) Int Cl.⁷: **G06F 17/30**

(21) Deutsches Aktenzeichen: **600 07 633.4**

(86) PCT-Aktenzeichen: **PCT/US00/18280**

(96) Europäisches Aktenzeichen: **00 945 118.8**

(87) PCT-Veröffentlichungs-Nr.: **WO 01/03010**

(86) PCT-Anmeldetag: **30.06.2000**

(87) Veröffentlichungstag

der PCT-Anmeldung: **11.01.2001**

(97) Erstveröffentlichung durch das EPA: **03.04.2002**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **07.01.2004**

(47) Veröffentlichungstag im Patentblatt: **18.11.2004**

(30) Unionspriorität:

346245 01.07.1999 US

(84) Benannte Vertragsstaaten:

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

(73) Patentinhaber:

Honeywell Inc., Morristown, N.J., US

(72) Erfinder:

FOSLIEN, Wendy, Minneapolis, US; HARP, A., Steven, Coon Rapids, US; LAKSHMINARAYAN, Kamakshi, Minneapolis, US; MYLARASWAMY, A., Dinkar, St. Anthony, US

(74) Vertreter:

derzeit kein Vertreter bestellt

(54) Bezeichnung: **INHALT-BASIERTE WIEDERGABE VON SERIENDATEN**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelebt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung

Technisches Gebiet

[0001] Die vorliegende Erfindung betrifft Datenserien und insbesondere das Abrufen von in großen Datensequenzen enthaltenen Daten.

Allgemeiner Stand der Technik

[0002] In vielen Industriezweigen verwendet man große Datenlager zum Verfolgen von Variablen über relative lange Spannen der Zeit oder des Raums hinweg. Mehrere Umgebungen, wie z. B. chemische Anlagen, Raffinerien und Gebäudesteuerung, verwenden als Prozeßvorgeschichten bekannte Datensätze zur Archivierung der Aktivität einer großen Anzahl von Variablen über die Zeit hinweg. Prozeßvorgeschichten verfolgen in der Regel Hunderte von Variablen und sind im wesentlichen Zeitserien mit hoher Dimensionalität. Die in Prozeßvorgeschichten enthaltenen Daten sind für vielfältige Zwecke nützlich, darunter z. B. Prozeßmodellaufbau, Optimierung, Steuersystemdiagnose und Analyse von Vorfällen (abnormalen Ereignissen).

[0003] Große Datensequenzen werden auch in anderen Gebieten verwendet, um die Aktivität von Variablen über die Zeit oder den Raum hinweg zu archivieren. Auf dem medizinischen Gebiet kann man durch Überwachen bestimmte biologischer Meßwerte, wie z. B. des Pulses, des Blutdrucks und dergleichen, wertvolle Einsichten gewinnen. Andere Gebiete sind z. B. Wirtschaft, Meterologie und Telemetrie.

[0004] In diesen und anderen Gebieten werden Ereignisse durch Datenmuster in einer oder mehreren der Variablen charakterisiert, wie z. B. eine schnelle Zunahme der Temperatur, mit der eine schnelle Zunahme des Drucks einhergeht. Es ist also wünschenswert, diese Datenmuster aus der Datensequenz als Ganzes zu extrahieren. Datensequenzen wurden üblicherweise durch Verwendung von Techniken wie z. B. Datenbankabfragesprachen analysiert. Solche Techniken ermöglichen es einem Benutzer, eine Datensequenz auf Prozeßvariablen von besonderem Interesse zugeordnete Daten abzufragen, können aber nicht angemessen auf der Zeit basierende Merkmale als Abfragekriterien integrieren. Außerdem können viele Datenmuster mit herkömmlichen Datenbankabfragesprachen nur schwer beschrieben werden. Darüber hinaus beeinträchtigt das Fehlen einer intuitiven Schnittstelle die Effizienz für viele Benutzer.

[0005] Um das Abfragen von Datensequenzen zu erleichtern, wurden sogenannte graphische Abfragesprachen entwickelt, die eine graphische Benutzeroberfläche (GUI) zur Eingabe von Standardbefehlen der Abfragesprache bieten. Aber auch bei Ver-

wendung dieser graphischen Abfragesprachen ist es jedoch schwierig, zeitliche Merkmal mengen oder Muster zu spezifizieren, die interessierende Ereignisse charakterisieren.

[0006] Ein weiteres Hindernis für die effiziente Analyse von Datensequenzen ist ihr Volumen. Da Datensequenzen viele Variablen über relativ große Zeitspannen hinweg verfolgen, sind sie in der Regel sowohl breit als auch tief. Folglich liegt die Größe einiger Datensequenzen in der Größenordnung von Gigabytes. Außerdem sind die meisten aufgezeichneten Daten häufig irrelevant. Aufgrund dieser Schwierigkeiten sind bestehende Techniken zum Extrahieren von Datenmustern aus Datensequenzen sowohl zeitaufwendig als auch umständlich.

[0007] Aus EP-A-0742525 ist ein Verfahren zum Finden von Datenmustern in einer Datensequenz bekannt. Aus US-A-5799301 ist ein Datenmusterauffinderverfahren, das Schablonen verwendet, bekannt.

[0008] Gemäß der vorliegenden Erfindung wird ein Verfahren zum Finden übereinstimmender Datenserienmuster in einer Datenseriensequenz, die ein Ähnlichkeitskriterium in bezug auf ein Ziellatenserienmuster erfüllen, bereitgestellt, das die folgenden Schritte umfaßt:

Anzeigen einer graphischen Darstellung mindestens eines Teils der Datenseriensequenz unter Verwendung einer graphischen Benutzeroberfläche; Empfangen einer Eingabe von einem Benutzer, die das Ziellatenserienmuster in der Datensequenz und das Ähnlichkeitskriterium definiert, unter Verwendung der graphischen Benutzeroberfläche; und Anwenden einer Mustererkennung auf die Datenseriensequenz, um die übereinstimmenden Datenserienmuster zu finden, die das Ähnlichkeitskriterium in bezug auf das Ziellatenserienmuster erfüllen.

[0009] Gemäß einem weiteren Aspekt wird eine Computereinrichtung zur Verwendung beim Finden übereinstimmender Datenserienmuster in einer Datenseriensequenz, die ein Ähnlichkeitskriterium in bezug auf ein Ziellatenserienmuster erfüllen, bereitgestellt, wobei die Einrichtung für folgendes konfiguriert ist:

Anzeigen einer graphischen Darstellung mindestens eines Teils einer Datenseriensequenz unter Verwendung einer graphischen Benutzeroberfläche; Empfangen von Daten von einem Benutzer, um das Ziellatenserienmuster in der Datensequenz und ein Ähnlichkeitskriterium zu definieren, unter Verwendung der graphischen Benutzeroberfläche; und Anwenden eines Mustererkennungsalgorithmus auf die Datenseriensequenz, die das Ähnlichkeitskriterium in bezug auf das Ziellatenserienmuster erfüllt.

Kurze Beschreibung der Zeichnungen

[0010] **Fig.** 1 ist ein Systemdiagramm einer Computeranordnung zum Analysieren einer Prozeßvorgeschichte gemäß einem Ausführungsbeispiel der vorliegenden Erfindung.

[0011] **Fig.** 2 ist ein Flußdiagramm eines Verfahrens zum Analysieren einer Datensequenz gemäß einem weiteren Ausführungsbeispiel der vorliegenden Erfindung.

[0012] **Fig.** 3 zeigt die Aktivität von fünf interessierenden Prozeßvariablen gemäß einem konkreten Anwendungsbeispiel für eine Ausführungsform der vorliegenden Erfindung.

[0013] **Fig.** 4 zeigt die Aktivität der fünf in **Fig.** 3 dargestellten Prozeßvariablen und außerdem die Benutzerauswahl von Bereichen von besonderem Interesse in zwei der Prozeßvariablen.

[0014] **Fig.** 5, 6 und 7 zeigen eine Dialogbox mit Reitern zum Angeben von Suchkriterien zum Finden von Übereinstimmungen gemäß einer Ausführungsform der vorliegenden Erfindung.

[0015] **Fig.** 8 zeigt eine Menge von gefundenen Übereinstimmungen gemäß einem Operationsbeispiel einer Ausführungsform der vorliegenden Erfindung.

[0016] **Fig.** 9 und 10 zeigen eine Dialogbox mit Reitern zur Untersuchung einer bestimmten gefundenen Übereinstimmung gemäß einer weiteren konkreten Ausführungsform der vorliegenden Erfindung.

[0017] **Fig.** 11 ist ein Flußdiagramm einer beispielhaften Technik zum Durchsuchen einer Datensequenz auf bestimmte interessierende Muster gemäß einer weiteren Ausführungsform der vorliegenden Erfindung.

[0018] **Fig.** 12 zeigt die Benutzerdefinition eines Suchziels, das aus zusammenfallenden Datenmustern in zwei Variablen besteht.

[0019] **Fig.** 13 zeigt eine beispielhafte Dialogbox zum Definieren eines Suchziels durch Auswahl einer Datenmusterschablone aus einer Palette von Schablonen.

Ausführliche Beschreibung

[0020] Die vorliegende Erfindung ist auf in der Industrie als Datensequenzen bekannte Ansammlungen von Daten anwendbar. Die Erfindung hat sich als besonders vorteilhaft für die Verwendung bei Datensequenzen erwiesen, bei denen Ereignisse von besonderem Interesse durch relativ kurze Datense-

quenzen dargestellt werden, die in großen Datensätzen enthalten sind. Verschiedene Aspekte der Erfindung können durch eine Besprechung verschiedener in solchen Umgebungen wirkender Anwendungsbeispiele verständlich werden.

[0021] Gemäß einem Ausführungsbeispiel der vorliegenden Erfindung wird eine Datensequenz durch Verwendung visuell orientierter Musterübereinstimmungs- und Merkmalanalysetechniken analysiert. Eine graphische Benutzeroberfläche (GUI) erleichtert das Angeben von interessierenden Ereignissen, die in der Datensequenz gefunden werden sollen. Der Benutzer verwendet die GUI zur Übersetzung der in dem Ereignis enthaltenen quantitativen Informationen in eine Beschreibung, die von einem Suchalgorithmus zum Finden von Übereinstimmungen in der Datensequenz verwendet wird. Die Übersetzung wird durch Auswahl eines Datensegments auf einem Graphen oder durch Verwendung einer Schablone, die die interessierenden Daten zum Beispiel als abklinende Sinusfunktion beschreibt, durchgeführt.

[0022] Nachdem ein Ereignis spezifiziert wurde, wird eine beliebige von mehreren Mustererkennungstechniken verwendet, um die Datensequenz nach dem Ereignis ähnlichen Sequenzen zu durchsuchen. Die Übereinstimmungen werden dem Benutzer als eine eingestufte Auflistung präsentiert, der die Übereinstimmungsergebnisse untersuchen und diejenigen auswählen kann, die am besten zu seinen Bedürfnissen passen, anstatt alle Daten in der Datensequenz zu scannen. Der Benutzer kann also interessierende Teile der Datensequenz schnell und leicht ansehen, ohne visuell eine große Menge uninteressanter Daten durchsehen zu müssen. Außerdem hat der Benutzer den Vorteil, daß er die Suchkriterien auf eine mit der Beschaffenheit der Daten kompatible Weise spezifizieren kann, da zeitabhängige Plots oder Schablonen ein natürliches Verfahren zum Definieren interessierender zeitabhängiger Muster sind.

[0023] Nunmehr mit Bezug auf die Zeichnungen zeigt **Fig.** 1 eine beispielhafte Computeranordnung **100** zur Analyse einer Datensequenz. Diese Computeranordnung **100** enthält eine Vielzweckdatenverarbeitungseinrichtung, wie z. B. einen Computer **102**. Der Computer **102** enthält eine Verarbeitungseinheit **104**, einen Speicher **106** und einen Systembus **108**, der die verschiedenen Systemkomponenten wirksam an die Verarbeitungseinheit **104** ankoppelt. Eine oder mehrere Verarbeitungseinheiten **104** wirken als eine einzige Zentralverarbeitungseinheit (CPU) oder als eine parallele Verarbeitungsumgebung.

[0024] Die Computeranordnung **100** enthält weiterhin eines oder mehrere Datenspeichergeräte zum Speichern und Lesen von Programm- und anderen Daten. Zu Beispielen für solche Datenspeichergeräte gehören ein Festplattenlaufwerk **110** zum Lesen und

Beschreiben einer (nicht gezeigten) Festplatte, ein Magnetplattenlaufwerk **112** zum Lesen oder Beschreiben einer (nicht gezeigten) herausnehmbaren Magnetplatte und ein optisches Plattenlaufwerk **114** zum Lesen oder Beschreiben einer (nicht gezeigten) herausnehmbaren optischen Platte, wie z. B. einer CD-ROM oder eines anderen optischen Mediums.

[0025] Das Festplattenlaufwerk **110**, das Magnetplattenlaufwerk **112** und das optische Plattenlaufwerk **114** sind durch eine Festplattenlaufwerkschnittstelle **116**, eine Magnetplattenlaufwerkschnittstelle **118** bzw. eine Schnittstelle **120** für ein optisches Plattenlaufwerk mit dem Systembus **108** verbunden. Diese Laufwerke und ihre zugeordneten computerlesbaren Medien liefern eine nichtflüchtige Speicherung von computerlesbaren Anweisungen, Datenstrukturen, Programmodulen und anderen Daten zur Verwendung durch die Computeranordnung **100**. In Verbindung mit der vorliegenden Erfindung kann jede beliebige Art von computerlesbaren Medien, die Daten, die für einen Computer zugänglich sind, speichern können, verwendet werden, wie z. B. Magnetkassetten, Flash-Speicherarten, DVDs (Digital Versatile Discs), Bernoulli-Kassetten, Direktzugriffsspeicher (RAMs) und Nurlesespeicher (ROMs).

[0026] In einem maschinenlesbaren Medium, wie z. B. der Festplatte, magnetischen Platte, optischen Platte, ROM, RAM oder einem elektrischen Signal, wie z. B. einem durch einen Kommunikationskanal empfangenen elektronischen Datenstrom, kann eine Anzahl von Programmodulen gespeichert oder codiert werden. Zu diesen Programmodulen gehören ein Betriebssystem, ein oder mehrere Anwendungsprogramme, andere Programmodule und Programmdaten.

[0027] Durch einen Adapter **124** oder eine andere Schnittstelle ist ein Monitor **122** mit dem Systembus **108** verbunden. Zusätzlich kann die Computeranordnung **100** andere (nicht gezeigte) Peripherieausgabegeräte enthalten, wie z. B. Lautsprecher und Drucker.

[0028] Die Computeranordnung **100** kann unter Verwendung logischer Verbindungen mit einem oder mehreren (nicht gezeigten) abgesetzten Computern in einer vernetzten Umgebung wirken. Diese logischen Verbindungen werden unter Verwendung eines Kommunikationsgeräts implementiert, das an die Computeranordnung **100** angekoppelt oder integraler Bestandteil dieser ist. Die zu analysierende Datensequenz kann in der vernetzten Umgebung auf einem abgesetzten Computer verankert sein. Der abgesetzte Computer kann ein anderer Computer, ein Server, ein Router, ein Netzwerk-PC, ein Client oder eine Peer-Einrichtung oder ein anderer gemeinsamer Netzwerkknoten sein. **Fig. 1** zeigt die logische Verbindung als eine Netzwerkverbindung **126**, die durch

eine Netzwerkschnittstelle **128** eine Schnittstelle mit der Computeranordnung **100** aufweist. Solche Vernetzungsumgebungen sind in Büronetzwerken, firmenweiten Computernetzwerken, Intranetzen und dem Internet, also allen Arten von Netzwerken, überall anzutreffen. Für Fachleute ist erkennbar, daß die gezeigten Netzwerkverbindungen als Beispiel angegeben sind und daß andere Mittel und Kommunikationseinrichtungen zur Herstellung einer Kommunikationsverbindung zwischen den Computern verwendet werden können.

[0029] **Fig. 2** zeigt ein beispielhaftes Verfahren **200** zum Durchsuchen einer Datensequenz nach einem Ereignis von besonderem Interesse gemäß einer Ausführungsform der vorliegenden Erfindung. Als erstes wählt der Benutzer in einem Block **202** eine Datenquelle, wie z. B. eine Datei, eine Vorgeschichtedatenbank oder eine CD-ROM. Die Datei kann in einem Standard-Dateiformat, z. B. ASCII-Text, oder in einem anwendungsspezifischen Format gespeichert sein.

[0030] Wie in einem Block **204** abgebildet, wählt der Benutzer als nächstes eine Prozeßvariable, die dann als Funktion der Zeit in einem Graphen aufgetragen wird. Es sollte beachtet werden, daß der Benutzer mehr als eine Prozeßvariable zur Betrachtung wählen kann. **Fig. 3** zeigt ein konkretes Anwendungsbeispiel, bei den fünf Prozeßvariablen zur Betrachtung in einem Anzeigebereich **302** einer GUI **300** gewählt wurden. Wenn der Benutzer eine Prozeßvariable auswählt, indem er mit der rechten Maustaste darauf klickt, erscheint eine Dialogbox, die es dem Benutzer ermöglicht, Eigenschaften der Prozeßvariablen, wie z. B. die Datei, in der ihre Daten gespeichert sind, die Abtastperiode, Anzahl von Abtastwerten und Dauer, für die die Prozeßvariable überwacht wurde, betrachten kann.

[0031] Die gewählten Prozeßvariablen sind als Reiter **304** in einer hierarchischen Baumstruktur **306** angegeben, die in einem Teil **308** der GUI **300** angezeigt wird. In dem Anzeigebereich **302** wird jede gewählte Prozeßvariable mit einem Graphen **301** angezeigt. Eine Werkzeugeiste **312** enthält Skalierungssteuerungen **314**, um eine schnelle Auswahl der Zeitmaßstäbe zur Betrachtung der gewählten Prozeßvariablen zu ermöglichen. Vergrößerungstasten **316** erleichtern das Herein- und Herauszoomen in und aus interessierenden Bereichen. Zusätzlich ermöglichen es die Rolltasten **318** dem Benutzer, seiten- oder grid-weise durch die Graphen **310** zu rollen. Außerdem kann der Benutzer die Rolltasten **318** benutzen, um zum Anfang oder zum Ende der verfügbaren Daten zu springen.

[0032] Ein Rollbalken **320** gibt den Teil der Datensequenz an, den der Benutzer betrachtet. Wenn der Benutzer z. B. den Anfang der Datensequenz betrach-

tet, erscheint der Rollbalken **320** auf der linken Seite des Anzeigebereichs **302**. Wenn der Benutzer dagegen das Ende der Datensequenz betrachtet, erscheint der Rollbalken **320** auf der rechten Seite. Der Benutzer kann nicht nur ausmachen, welchen Teil der Datensequenz er betrachtet, sondern auch auf den Rollbalken **320** klicken, um zu einer bestimmten Stelle in der Datensequenz zu springen.

[0033] Nachdem der Benutzer die Prozeßvariable der -variablen ausgewählt hat, schreitet der Fluß zu einem Block **206** voran, in dem der Benutzer eine oder mehrere interessierende Zielregionen definiert, die gewöhnlich Ereignisse sind. Der Benutzer kann interessierende Ereignisse auf mehrere Weisen spezifizieren. Z. B. kann der Benutzer aus einer vordefinierten Palette von Trendmustern einen Prototyp auswählen. **Fig. 13** zeigt eine beispielhafte Dialogbox **1300**, die zur Auswahl eines Prototyps oder Exemplars aus einer Palette verwendet wird. Diese Palette enthält mehrere vordefinierte Trendmuster **1302**, **1304**, **1306**, **1308** und **1310**. Zusätzlich können außerdem eines oder mehrere Benutzerdefinierte Trendmuster **1312** in der Palette dargestellt werden.

[0034] Wenn das Ereignis durch keinen der Prototypen in der Palette angemessen dargestellt wird, kann der Benutzer das Ereignis z. B. durch Ziehen des Cursors über den interessierenden Bereich definieren. **Fig. 4** zeigt die Spezifikation zweier Ereignisse **402** auf diese Weise. Während Zielergebnisse in dem Anzeigebereich **302** definiert werden, zeigt die hierarchische Baumstruktur **306** die definierten Ereignisse als Ziele **404** in einer untergeordneten Beziehung mit entsprechenden Reitern **304**.

[0035] Gemäß einer konkreten Ausführungsform können Ereignisse durch in mehreren Prozeßvariablen auftretende Muster charakterisiert werden. Z. B. kann der Benutzer spezifizieren, daß ein Ereignis in einer schnellen Zunahme der Temperatur mit damit einhergehender schneller Abnahme des Drucks besteht. Dieses Ereignis würde durch zwei Prozeßvariablen (Temperatur und Druck) definiert. Der Benutzer kann spezifizieren, wie nahe in bezug auf die Zeit diese Phänomene auftreten müssen. **Fig. 12** zeigt eine beispielhafte Suchzieldefinition **1200**, die zwei zuvor definierte einfache Merkmale **1202** und **1204** enthält. Eine Verknüpfung **1206** verbindet die einfachen Merkmale **1202** und **1204**, wodurch angezeigt wird, daß die beiden einfachen Merkmale **1202** und **1204** zusammen auftreten müssen, um als eine Übereinstimmung in Frage zu kommen. Insbesondere wird durch alle paarweisen Übereinstimmungen mit den beiden einfachen Merkmalen **1202** und **1204** unter Berücksichtigung zeitlicher Einschränkungen eine Übereinstimmung mit der Suchzieldefinition **1200** festgelegt. In dem in **Fig. 12** abgebildeten spezifischen Beispiel gelten die zeitlichen Einschränkungen für die Zeit zwischen dem Ende des einfachen Merk-

mals **1202** und dem Anfang des einfachen Merkmals **1204**. Zeitliche Einschränkungen können als Alternative über die Zeit zwischen den Anfängen beider einfacher Merkmale **1202** und **1204** oder ihren Enden definiert werden. Zusätzlich können relative Zeitgrenzen, analog zu den nachfolgend in Verbindung mit **Fig. 6** besprochenen Kompressions- und Expandiergrenzen, definiert werden. Als Alternative können zeitliche Einschränkungen absolut definiert werden, wie z. B. als eine gegebene Anzahl von Stunden.

[0036] Es ist ohne weiteres erkennbar, daß die Suchzieldefinition **1200** einer logischen AND-Operation analog ist. Mehrvariablensuchziele müssen nicht konjunktiv definiert werden. Sie können auch disjunktiv oder negativ, d. h. analog einer logischen OR- oder NOT-Operation, definiert werden. Diese Arten von Zieldefinitionen können auch zu komplexen Fragen kombiniert werden, z. B. „Merkmal A und entweder Merkmal B oder Merkmal C, aber nicht Merkmal D“. Zeitliche Einschränkungen können für beliebige dieser Arten von Zieldefinitionen spezifiziert werden.

[0037] Nachdem das Ziel bzw. die Ziele definiert worden sind, benutzt als nächstes der Benutzer die GUI **300** zum Suchen nach Regionen der Datensequenz, die angemessen mit dem Ziel bzw. den Zielen übereinstimmen, wie in einem Block **208** abgebildet. Als erstes definiert der Benutzer Suchkriterien, wie z. B. wie gut eine Region der Datensequenz mit einem Ziel übereinstimmen muß, um als eine „Übereinstimmung“ betrachtet zu werden. **Fig. 5–7** zeigen eine beispielhafte Dialogbox **500** zum Spezifizieren von Suchkriterien. In der in **Fig. 5** dargestellten Ansicht wurde ein Amplituden-Reiter **502** gewählt. Die Schieber **504** und **506** ermöglichen es dem Benutzer, die zulässigen relativen Amplitudengrenzen für Übereinstimmungen, ausgedrückt als Schrumpf- bzw. Wachstumsgrenzen, einzustellen. Wenn der Benutzer z. B. eine Wachstumsgrenze von Zwei spezifiziert, kann eine Übereinstimmung höchstens zweimal die Effektivwertamplitude (RMS) des Ziels aufweisen. Umgekehrt bedeutet eine Schrumpfgrenze von Zwei, daß eine Übereinstimmung mindestens die Hälfte der RMS-Amplitude des Ziels aufweisen muß.

[0038] **Fig. 6** zeigt die Dialogbox **500** mit gewähltem Zeitverzerrungsreiter **602**. Unter diesem Reiter ermöglichen es die Schieber **604** und **606** dem Benutzer, die zulässigen relativen Dauergrenzen für Übereinstimmungen, ausgedrückt als Kompressions- bzw. Expandierungsgrenzen, einzustellen. Eine Kompressionsgrenze von Zwei bedeutet z. B., daß eine Übereinstimmung mindestens die Hälfte der Dauer des Ziels aufweisen muß. Ähnlich bedeutet eine Expandierungsgrenze von Zwei, daß eine Übereinstimmung bis zu zweimal die Dauer des Ziels aufweisen kann. Eine Auflösungssteuerung **608** ermöglicht es dem Benutzer, einen Grad des Kompromisses zwischen der Menge an rechnerischen Bemühun-

gen, die bei der Suche aufgebracht wurde, und der Genauigkeit der Einstufung der Ergebnisse zu spezifizieren. Eine hohe Auflösung bewirkt z. B., daß der Suchalgorithmus mehr Mühe, d. h. mehr Zeit, für das Finden von Übereinstimmungen zuteilt, führt aber zu einer genaueren Einstufung.

[0039] Die oben in Verbindung mit **Fig. 5** und **6** besprochenen relativen Amplituden- und Dauergrenzen können harte Einschränkungen sein, die in Frage kommende Muster, die ansonsten mit dem Suchziel übereinstimmen würden, heraussortieren, sich aber anderweitig nicht auf die Qualität der Anpassung oder wahrgenommenen Ähnlichkeit auswirken. Als Alternative können diese Grenzen als Soft-Einschränkungen definiert werden, und in diesem Fall wirken sich Abweichungen von der Amplitude oder Dauer des Suchziels tatsächlich negativ auf die Anpassungsqualität aus.

[0040] **Fig. 7** zeigt die Dialogbox **500** mit gewähltem Such-Reiter **702**. Unter diesem Reiter ermöglicht ein Übereinstimmungsschwellenschieber **704** dem Benutzer die Spezifikation einer Übereinstimmungsschwelle, die beschreibt, wie gut ein Teil der Datensequenz mit dem Ziel übereinstimmen muß, um als ein Treffer in Frage zu kommen. Die Übereinstimmungsschwelle hat einen Wert zwischen Null und Eins. Eine Übereinstimmungsschwelle von Eins erfordert eine perfekte Übereinstimmung, während eine Übereinstimmungsschwelle von Null überhaupt keine Beziehung mit dem Ziel erfordert. Die Übereinstimmungsschwelle kann auch negativ sein, wodurch Regionen mit dem Ziel entgegenlaufenden Trends angezeigt werden, wie z. B. ein fallender Trend, wenn das Ziel einen steigenden Trend enthält.

[0041] Eine Abwärtsabtastverhältnissesteuerung **706** ermöglicht es dem Benutzer, ein Abwärtsabtastverhältnis zu spezifizieren, das den Faktor bestimmt, um den die ursprüngliche Abtastrate der Datensequenz reduziert wird. Wenn man z. B. einen Trend betrachtet, der sich in einer Datensequenz, die in einmütigen Intervallen abgetastet wird, über drei Tage hinweg entwickelt, würden die drei Tage 4320 einminütige Abtastwerte enthalten. Es ist im allgemeinen Verschwendug, alle diese Abtastwerte zu betrachten. Der Benutzer kann also ein Abwärtsabtastverhältnis von z. B. sechzig wählen, um das Signal auf eine weniger häufige, aber immer noch effektive Rate, wie z. B. stündliche Abtastwerte, umabzutasten. Im allgemeinen wird das Abwärtsabtastverhältnis automatisch vom System eingestellt und erfordert keinen Eingriff durch den Benutzer.

[0042] Unter Verwendung der oben in Verbindung mit **Fig. 5–7** besprochenen Dialogboxen kann der Benutzer also Ähnlichkeitskriterien definieren, um den Bedürfnissen der konkreten Anwendung zu entsprechen. Zu diesen Ähnlichkeitskriterien können

Einschränkungen in bezug auf Zeit, Amplitude, Anpassungsqualität und andere Betrachtungen gehören.

[0043] Nachdem der Benutzer die Suchkriterien in Verbindung mit Block **208** spezifiziert hat, durchsucht das System die Datensequenz nach Übereinstimmungen, wie in einem Block **210** von **Fig. 2** abgebildet. Zum Finden der Übereinstimmungen kann ein beliebiger von vielfältigen Mustererkennungsalgorithmen verwendet werden. Gemäß einer Ausführungsform gibt das System dem Benutzer eine Auswahl mehrerer solcher Algorithmen. Diese Algorithmen verwenden denselben Rahmen zum Erhalten von Daten, Auswählen von Suchoptionen und zum Präsentieren von Ergebnissen. Intern können die Algorithmen jedoch sehr verschiedene Techniken zum Finden von Übereinstimmungen benutzen.

[0044] Es werden unten drei beispielhafte Techniken besprochen. Jede Technik nimmt eine Zielsequenz $T = \{t_1, t_2, \dots, t_k\}$ und einen Suchbereich $S = \{s_1, s_2, \dots, s_k\}$ an, woran Übereinstimmungen $M_k = \{k_0, k_1\}$ gefunden werden sollen. Die sich aus dem Suchalgorithmus ergebenden Übereinstimmungen sind insofern geschlossene Intervalle in S , als sie ein bestimmtes Ähnlichkeitskriterium μ erfüllen, das von dem konkreten Algorithmus abhängig sein kann.

[0045] Der Algorithmus empfängt außerdem als Eingabe Einschränkungen bezüglich der maximalen und minimalen Dauer (D^+ , D^-) akzeptabler Übereinstimmungen aus den Expandier- und Kompressionsgrenzen. Ähnlich werden aus den Wachstums- und Schrumpfgrenzen Einschränkungen bezüglich der maximalen und minimalen Amplitude (A^+ , A^-) akzeptabler Übereinstimmungen abgeleitet. Diese Einschränkungen werden relativ zu dem Ziel T ausgedrückt, z. B. wäre 1,0 identisch mit T , während 2,0 zweimal so lang oder intensiv wäre. Wahlweise können Algorithmen zusätzliche Parameter anbieten, um die Suche einzuschränken oder um einen Kompromiß zwischen Sorgfältigkeit und Geschwindigkeit der Suche zu spezifizieren.

[0046] **Fig. 11** ist ein Flußdiagramm einer beispielhaften Suchtechnik **1100**, die als Übereinstimmungsfiltertechnik mit Verzerrung bekannt ist. Als erstes wird wie in einem Block **1102** abgebildet eine mit Tiefpaß (LPF) gefilterte Version T_f des Ziels vorbereitet. Das Filtern des Ziels beseitigt Transienten, die für das interessierende Merkmal irrelevant sind. Die Grenzfrequenz des Tiefpaßfilters ist wahlweise benutzerwählbar, muß aber gegebenenfalls den Antialiasanforderungen für die Dezimierung genügen.

[0047] Für sehr lange Ziele oder Suchräume wird die LPF-Version T_f des Ziels um einen Faktor p dezimiert, um die Suche zu beschleunigen, wie in einem Block **1104** abgebildet. Die Dezimierung ergibt eine

dezimierte Zielsequenz $T_d = \{t_{d,0}, t_{d,1}, \dots, t_{d,K}\}$ der Länge $K' = K/p$. Es sollte beachtet werden, daß die Dezimierung für kürzere Ziele oder Suchräume nicht erforderlich ist.

[0048] Wie in einem Block **1106** abgebildet, werden als nächstes zeitlich verzerrte Versionen des Ziels T oder des dezimierten Ziels T_d vorbereitet. Diese zeitlich verzerrten Versionen werden als $W_0(T_d)$, $W_1(T_d)$, ..., $W_J(T_d)$ bezeichnet, wobei Verzerrungsoperatoren W_i verwendet werden. Bei einer Ausführungsform wird eine Spanne von linearen Expansions- und Kompressions-Operatoren, die den Bereich zulässiger Übereinstimmungsdauereinschränkungen (D^+ , D^-) abdecken, verwendet. Als Alternative können auch nichtlineare Verzerrungen verwendet werden.

[0049] Wie in einem Block **1108** dargestellt, wird der Suchbereich S dann genauso wie das Ziel T gefiltert und/oder dezimiert, wodurch ein transformierter Suchbereich S_d entsteht. Für relativ kleine Suchbereiche kann dieser Prozeß in einem einzigen Schritt durchgeführt werden. Für größere Suchbereiche kann der Prozeß in Schritten durchgeführt werden.

[0050] Als nächstes wird auf jedes zeitlich verzerrte Ziel in jeder Position I in S_d ein der Art des „gleitenden Fensters“ zugegriffen. Zur Berechnung des Grads der Übereinstimmung R_i für jede Position I wird ein Korrelationskoeffizient verwendet. Dieser Korrelationskoeffizient wird gewöhnlich als $\{\text{SIGMA } xy\} \text{ OVER } \text{SQRT } \{\text{SIGMA } x^2 \text{ SIGMA } y^2\}$ definiert, wobei die Terme „ x “ das standardisierte Ziel sind, mit einem Mittelwert von Null und einer Varianz von Eins, und wobei die Terme „ y “ die standardisierte potentielle Übereinstimmungsregion darstellen. Dieser Koeffizient stellt lose die Kovarianz der beiden Sequenzen, normiert durch ihre einzelnen Standardabweichungen, dar.

[0051] Dieses Maß reagiert nicht auf Amplitudendifferenzen zwischen dem Ziel und der potentiellen Übereinstimmung. Die Berechnung von R_i kann auf effiziente inkrementelle Weise als Teil einer Sequenz $\{R_0, R_1, \dots\}$ durchgeführt werden. Ein Block **1110** bildet diesen Prozeß ab.

[0052] Die Sammlung aller Positionen I mit $\{i: R_i > \mu\}$ enthält in Frage kommende Übereinstimmungen. In einem Block **1112** wird diese in Frage kommende Menge reduziert, indem Positionen zurückgewiesen werden, die eine wesentliche Überlappung auf einer beliebigen Position j mit einem höheren R_j aufweisen. In einem Block **1114** wird die in Frage kommende Menge weiter reduziert, indem Kandidaten beseitigt werden, die die Amplitudeneinschränkungen (A^+ , A^-) verletzen. Als letztes wird in einem Block **1116** diese reduzierte Menge von Kandidaten als die Menge von Übereinstimmungen M zurückgegeben.

[0053] Anstelle der oben in Verbindung mit **Fig. 11** besprochenen können andere Suchtechniken verwendet werden. Zum Beispiel wird eine als Spline-Übereinstimmungsfiltertechnik bekannte alternative Technik im wesentlichen ähnlich wie die Technik von **Fig. 11** durchgeführt. Bei der Spline-Übereinstimmungsfiltertechnik wird jedoch keine Tiefpaßfilterung durchgeführt. Statt dessen wird das Ziel T durch etwaige scheinbare Diskontinuitäten, wie z. B. Sprünge oder Flexionspunkte, partitioniert. Jedes kontinuierliche Segment wird dann durch den am besten passenden polynomischen Spline niedriger Ordnung ersetzt. Dieser Ersetzungsprozeß liefert eine idealisierte Version des Ziels zur Verwendung in den nachfolgenden Suchschritten.

[0054] Eine weitere Suchtechnik verwendet eine herkömmliche Transformationstechnik, die generisch als qualitative Trendanalyse (QTA) bekannt ist. Bei der QTA werden das ursprüngliche Ziel und der Suchbereich unter Verwendung eines zeitinvarianten Ansatzes für die Segmentierung der Datenmenge zu Ketten symbolischer Trendprimitiven zerlegt. Es kann eine beliebige von vielfältigen herkömmlichen Segmentierungstechniken verwendet werden, wie z. B. Wavelet-Zerlegung. Die Trendprimitiven oder Symbole sind gegenüber Amplituden- oder Daueränderungen invariant, obwohl die Maßstabsinformationen zur späteren Bezugnahme bewahrt werden. Mehrere Trendalphabete sind in der Technik bekannt. Ein solches Trendalphabet besteht aus geraden Linien und Kurven, die durch Kombinieren der geraden Linien gebildet werden. Nachdem der Suchbereich S und das Ziel T zu Ketten auf einem reduzierten Alphabet reduziert worden sind, wird der Suchprozeß zu einem Kettenvergleichsprozeß, für den verschiedene effiziente Algorithmen in der Technik bekannt sind. Die Anzahl übereinstimmender Stellen kann als ein Maß der Genauigkeit verwendet werden. Nachdem auf diese Weise eine Menge von Kandidaten bestimmt wurde, wird diese Menge durch Anwenden der Amplituden- und Dauereinschränkungen auf die behaltenen Maßstabsinformationen reduziert. Für Fachleute ist erkennbar, daß andere Suchtechniken als die hier besprochenen zum Finden von Übereinstimmungen in der Datensequenz verwendet werden können.

[0055] Nachdem der Suchalgorithmus Teile der Datensequenz, die angemessen mit dem Ziel übereinstimmen, gefunden hat, werden die Übereinstimmungen in dem Anzeigebereich **302** dem Benutzer präsentiert, wie in einem Block **212** von **Fig. 2** abgebildet. Zusätzlich können die Suchergebnisse zur weiteren Analyse zu anderen Anwendungen exportiert werden. Zum Beispiel können die Daten zu einem Vielzweckwerkzeug exportiert werden, das Frequenzinhalt in den den Suchergebnissen entsprechenden Daten berechnet.

[0056] **Fig. 8** zeigt die GUI **300**, so wie sie erscheint,

nachdem Übereinstimmungen für das Ziel T1-1, wie durch den schattierten Teil **802** angezeigt, gefunden wordensind. Die Übereinstimmungen werden durch die Übereinstimmungslabels **804** in einer untergeordneten Beziehung zu dem entsprechenden Ziel **404** in der hierarchischen Baumstruktur **306** angezeigt. Die Übereinstimmungslabel **804** werden in der Reihenfolge sortiert, wie gut die entsprechenden Übereinstimmungen auf das Zielereignis passen. Die den Übereinstimmungslabels **804** zugeordneten Symbole **806** geben die Übereinstimmungsqualität an. Bei der in **Fig. 8** gezeigten konkreten Implementierung zeigt ein voll schattiertes Karo eine sehr gute Anpassung an. Teilweise schattierte Karos zeigen etwas losere Anpassungen an.

[0057] Durch Klicken auf ein Symbol **806** kann der Benutzer eine bestimmte Übereinstimmung besuchen. Die Ansicht in dem Anzeigebereich **302** springt zu dem Teil der Datensequenz, der die Übereinstimmung enthält. Außerdem wird die in **Fig. 9** und **10** abgebildete Dialogbox **900** aufgerufen. Die Dialogbox **900** enthält zwei Reiter: einen Statistikreiter **902** und einen Graphen-Reiter **904**. Ein Anwählen des Statistikreiters **902** bewirkt die Anzeige der Ansicht von **Fig. 9**. Diese Ansicht enthält bestimmte Statistiken, die mit der Übereinstimmung zusammenhängen. Zum Beispiel wird bei der in **Fig. 9** abgebildeten konkreten Implementierung die Qualität der Anpassung angezeigt sowie die relative Dauer und Amplitude der Übereinstimmung, ausgedrückt als ein Teil der Dauer und Amplitude des Ziels. Außerdem werden der Bereich der Übereinstimmung in der Datensequenz und die Gesamtdauer der Übereinstimmung angezeigt.

[0058] Ein Anwählen des Graphen-Reiters **904** bewirkt dagegen eine Anzeige der Ansicht von **Fig. 10**. In dieser Ansicht kann der Benutzer die Übereinstimmung und das Ziel auf einem einzigen Graphen **1002** aufgetragen sehen. Der Benutzer kann also die Übereinstimmung und das Ziel visuell vergleichen. Eine Ankreuzbox **1004** bewirkt, wenn sie angekreuzt wird, daß die Dauer der Übereinstimmung auf die Dauer des Ziels in dem Graphen **1002** skaliert wird. Das Ankreuzen einer Ankreuzbox **1006** bewirkt, daß die Kurven der Übereinstimmung und des Ziels geglättet werden.

[0059] Nachdem der Benutzer die Übereinstimmungen zu seiner Befriedigung untersucht hat, kann der Benutzer das System nochmals abfragen, um eine weitere Suche durchzuführen, wie durch den Entscheidungsblock **2104** von **Fig. 2** angegeben. Wenn sich der Benutzer entscheidet, eine weitere Suche durchzuführen, kehrt der Fluß zu dem Block **206** zurück, und der Benutzer formuliert eine neue Suche.

[0060] Als Alternative kann der Benutzer die Suche verfeinern. Zum Beispiel kann der Benutzer die Ergebnisse auf Zeitsegmente einschränken, die län-

ger als eine spezifizierte Grenze sind, oder ein neues visuelles Merkmal spezifizieren, nach dem ausschließlich in den Suchergebnissen gesucht werden soll. Diese verfeinerte Suche wirkt dann an den Anfangssuchergebnissen, anstatt der gesamten Datensequenz.

[0061] Die Verwendung eines visuell orientierten Ansatzes für die Analyse von Datensequenzen ermöglicht dem Benutzer ein intuitives Spezifizieren von Ereignissen, ohne daß er traditionelle Abfragesprachen kennen muß. Zusätzlich kann der Benutzer mit Zuversicht vermeiden, große Mengen an für die aktuelle Abfrage nicht relevanten Daten zu untersuchen, und statt dessen interessierende Regionen schnell und leicht betrachten.

Patentansprüche

1. Verfahren zum Finden übereinstimmender Datenserienmuster in einer Datenseriensequenz, die ein Ähnlichkeitskriterium in bezug auf ein Zieldatenserienmuster erfüllen, mit den folgenden Schritten: Anzeigen einer graphischen Darstellung mindestens eines Teils der Datenseriensequenz unter Verwendung einer graphischen Benutzeroberfläche (**300**); Empfangen einer Eingabe von einem Benutzer, die das Zieldatenserienmuster (**400**) und das Ähnlichkeitskriterium definiert; Anwenden einer Mustererkennung auf die Datenseriensequenz, um die übereinstimmenden Datenserienmuster zu finden, die das Ähnlichkeitskriterium in bezug auf das Zieldatenserienmuster erfüllen, **dadurch gekennzeichnet**, daß das Zieldatenserienmuster in der Datenseriensequenz unter Verwendung der graphischen Benutzeroberfläche definiert wird.

2. Verfahren nach Anspruch 1, wobei das Definieren des Zieldatenserienmusters (**404**) das Auswählen mindestens eines Prototypmusters aus einer vordefinierten Palette von Mustern (**1302 ... 1310**) umfaßt.

3. Verfahren nach Anspruch 1, weiterhin mit den folgenden Schritten: Erstellen mehrerer zeitlich verzerrter Versionen mindestens eines Teils des Zieldatenserienmusters; Vergleichen mindestens einer der zeitlich verzerrten Versionen mindestens eines Teils des Zieldatenserienmusters (**404**); Vergleichen mindestens einer der zeitlich verzerrten Versionen mit mindestens einem Teil der Datenseriensequenz, um mehrere in Frage kommende Datenserienmuster in der Datensequenz zu bestimmen, die eine Übereinstimmungsschwelle in bezug auf die verglichene mindestens eine zeitlich verzerrte Version erfüllen; und Zurückweisen von in Frage kommenden Datenserienmustern, die Amplitudengrenzen verletzen.

4. Computereinrichtung (**100**) zur Verwendung beim Finden übereinstimmender Datenserienmuster in einer Datensiensequenz, die ein Ähnlichkeitskriterium in bezug auf ein Zieldatenserienmuster (**404**) erfüllen, wobei die Einrichtung für folgendes konfiguriert ist:

Anzeigen einer graphischen Darstellung mindestens eines Teils einer Datensiensequenz unter Verwendung einer graphischen Benutzeroberfläche (**300**); Empfangen von Daten von einem Benutzer, um das Zieldatenserienmuster und ein Ähnlichkeitskriterium zu definieren;

Anwenden eines Mustererkennungsalgorithmus auf die Datensiensequenz, die das Ähnlichkeitskriterium in bezug auf das Zieldatenserienmuster erfüllt, dadurch gekennzeichnet, daß die graphische Benutzeroberfläche so konfiguriert ist, daß der Benutzer das Zieldatenserienmuster in der Datensiensequenz definieren kann.

5. Einrichtung (**100**) nach Anspruch 4, die weiterhin für folgendes ausgelegt ist:

Erstellen mehrerer zeitlich verzerrter Versionen mindestens eines Teils des Zieldatenserienmusters (**404**);

Vergleichen mindestens einer der zeitlich verzerrten Versionen mit mindestens einem Teil der Datensiensequenz, um mehrere in Frage kommende Datenmuster in der Datensiensequenz zu bestimmen, die eine Übereinstimmungsschwelle in bezug auf die verglichene mindestens eine zeitlich verzerrte Version erfüllen; und

Zurückweisen von in Frage kommenden Datensiensemustern, die Amplitudengrenzen verletzen.

Es folgen 12 Blatt Zeichnungen

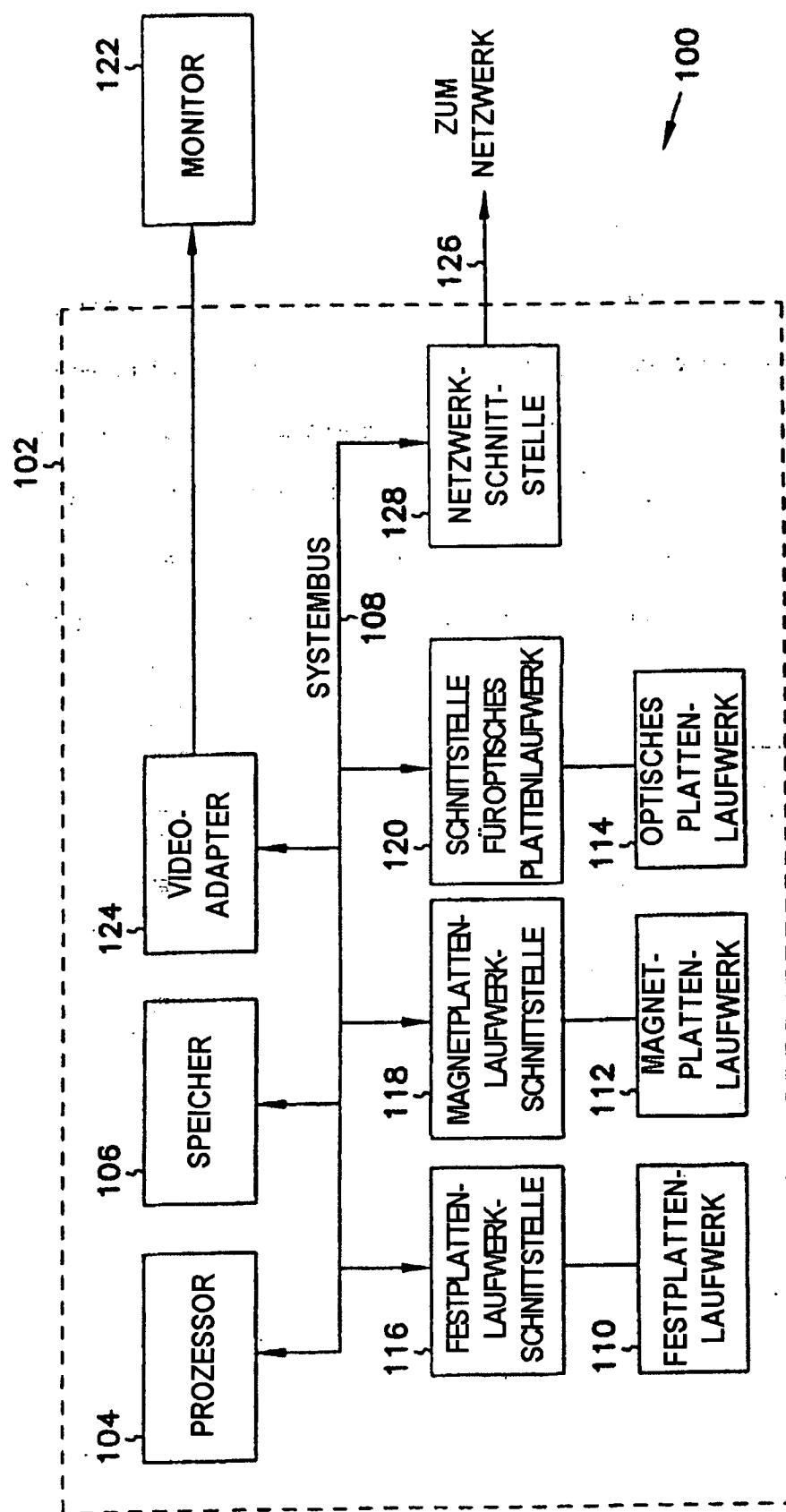


FIG. 1

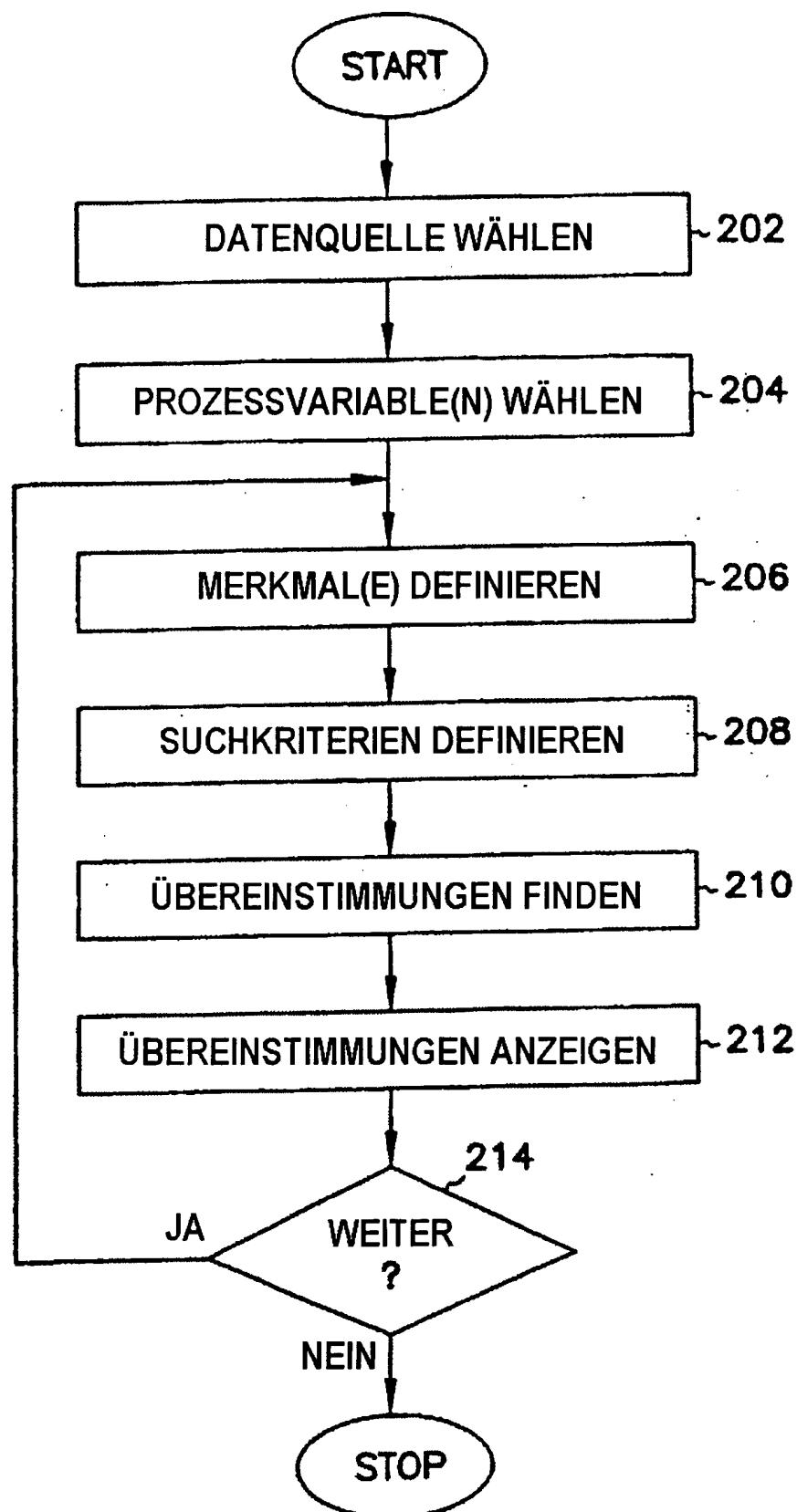


FIG. 2

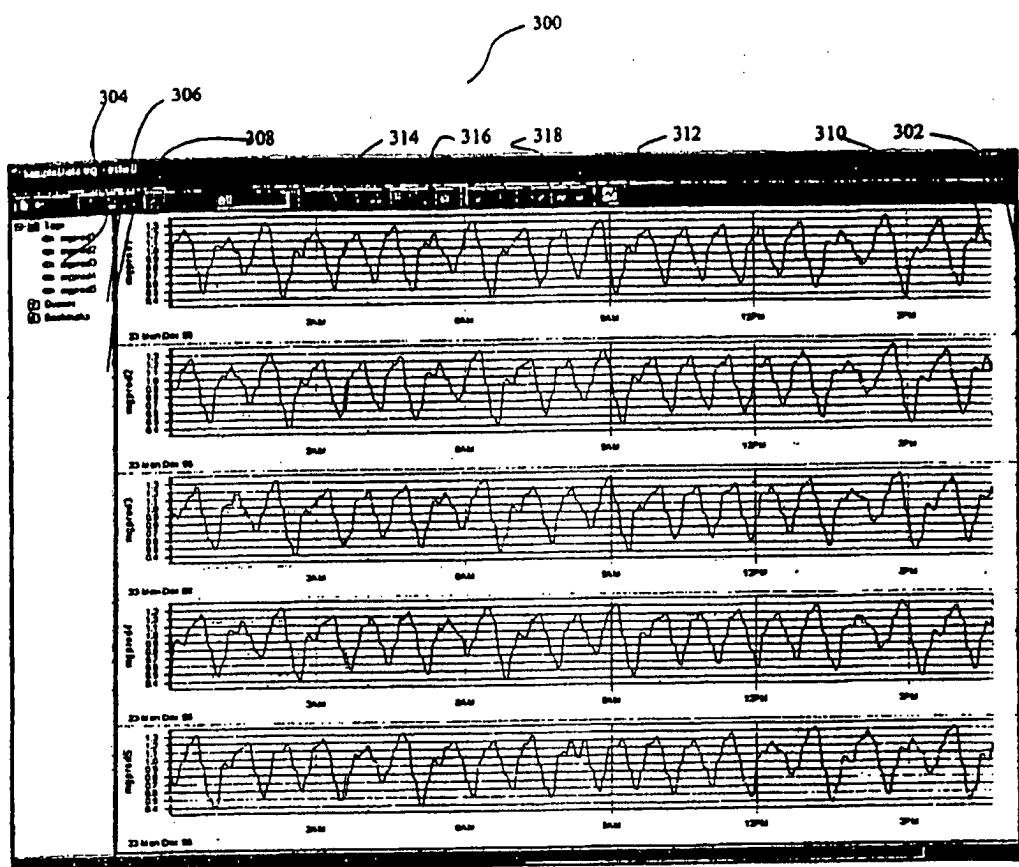


Fig. 3

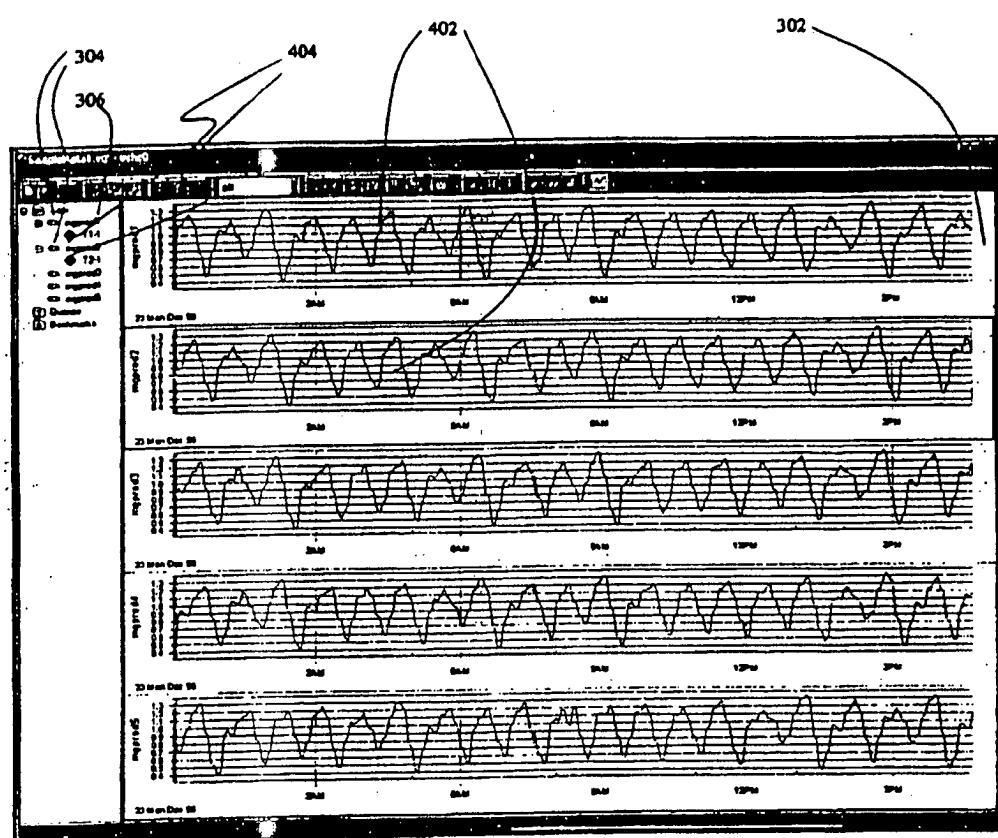


Fig. 4

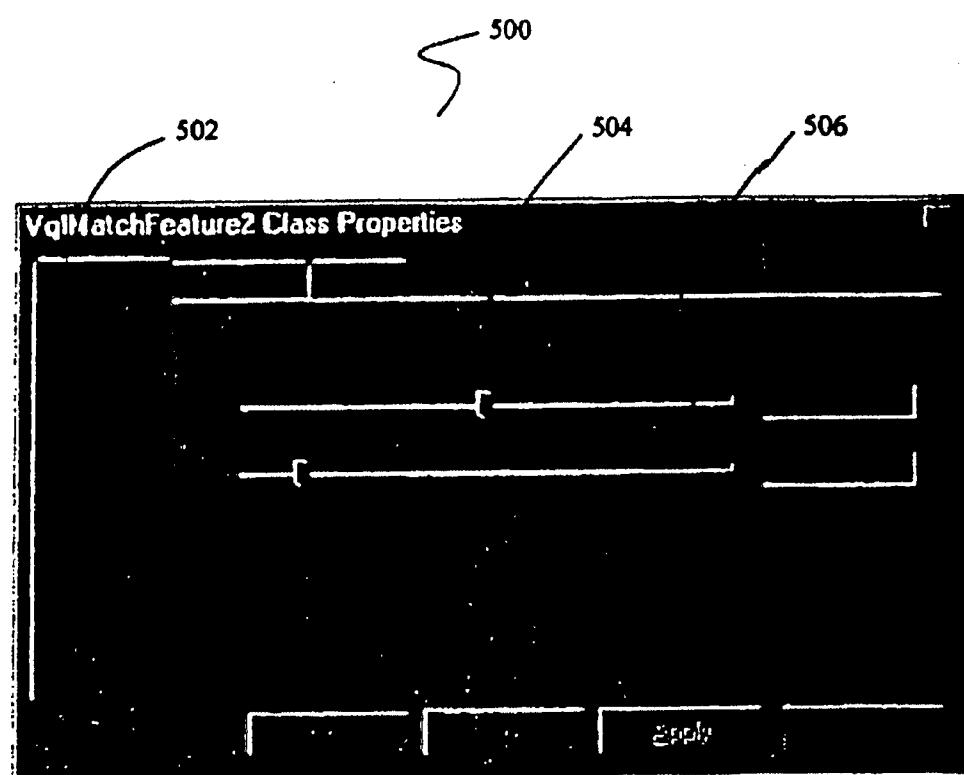


Fig. 5

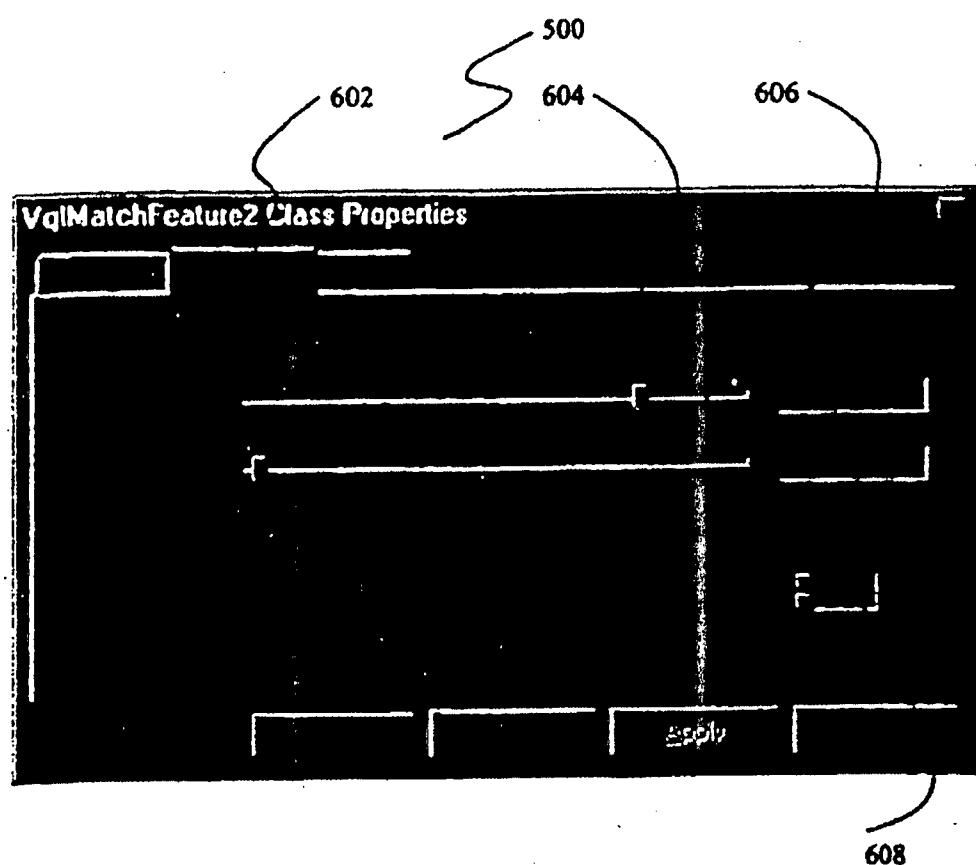


Fig. 6

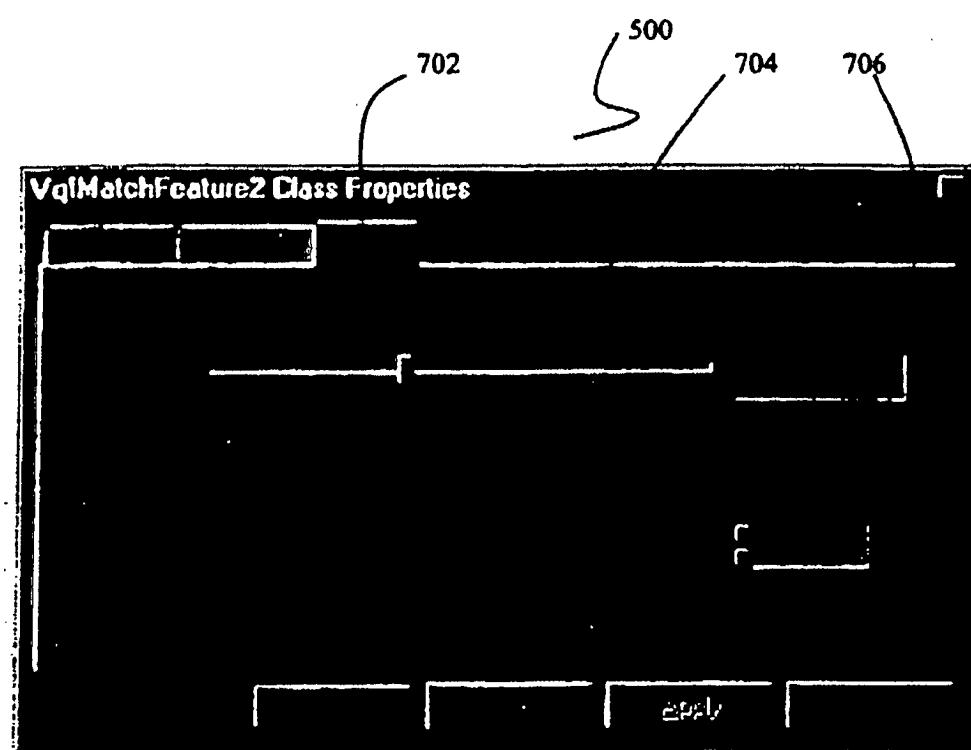


Fig. 7

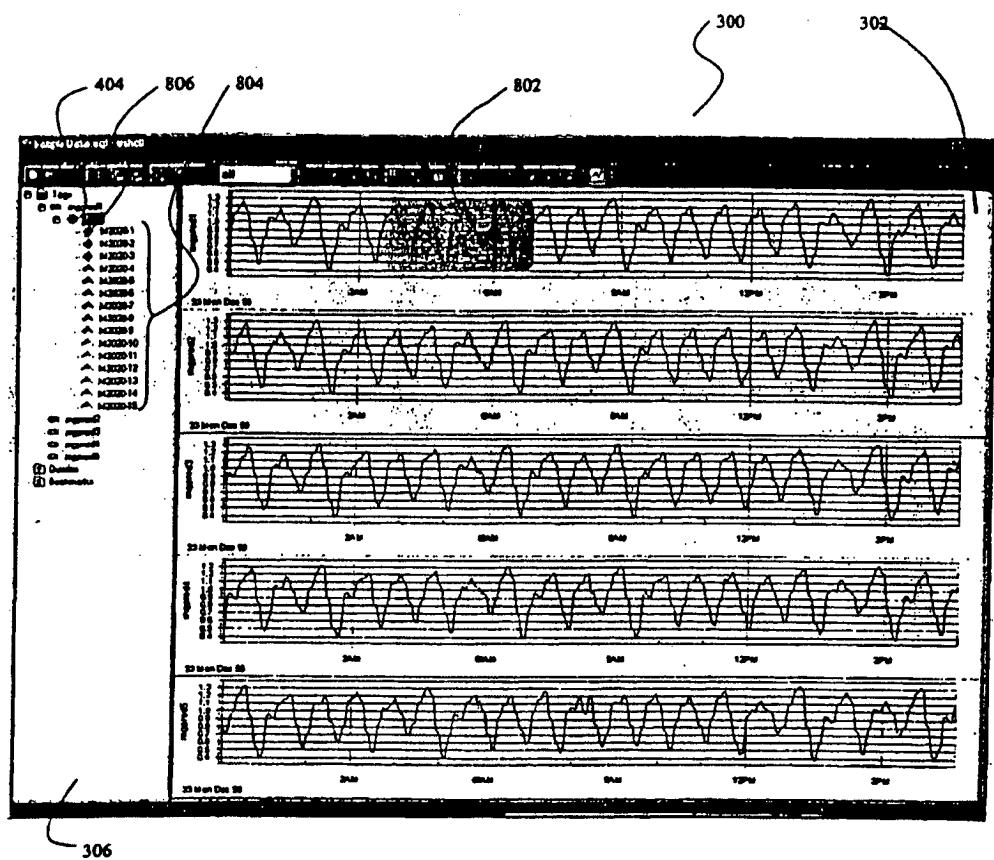


Fig. 8

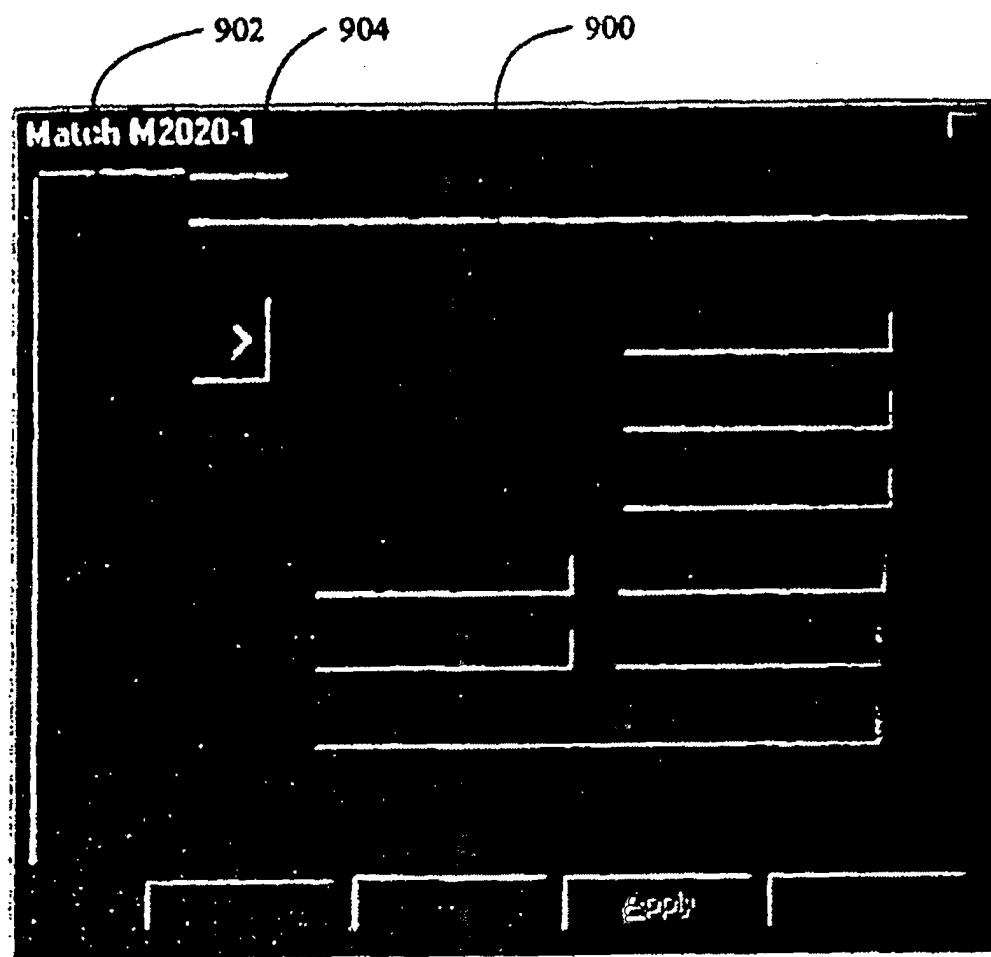


Fig. 9

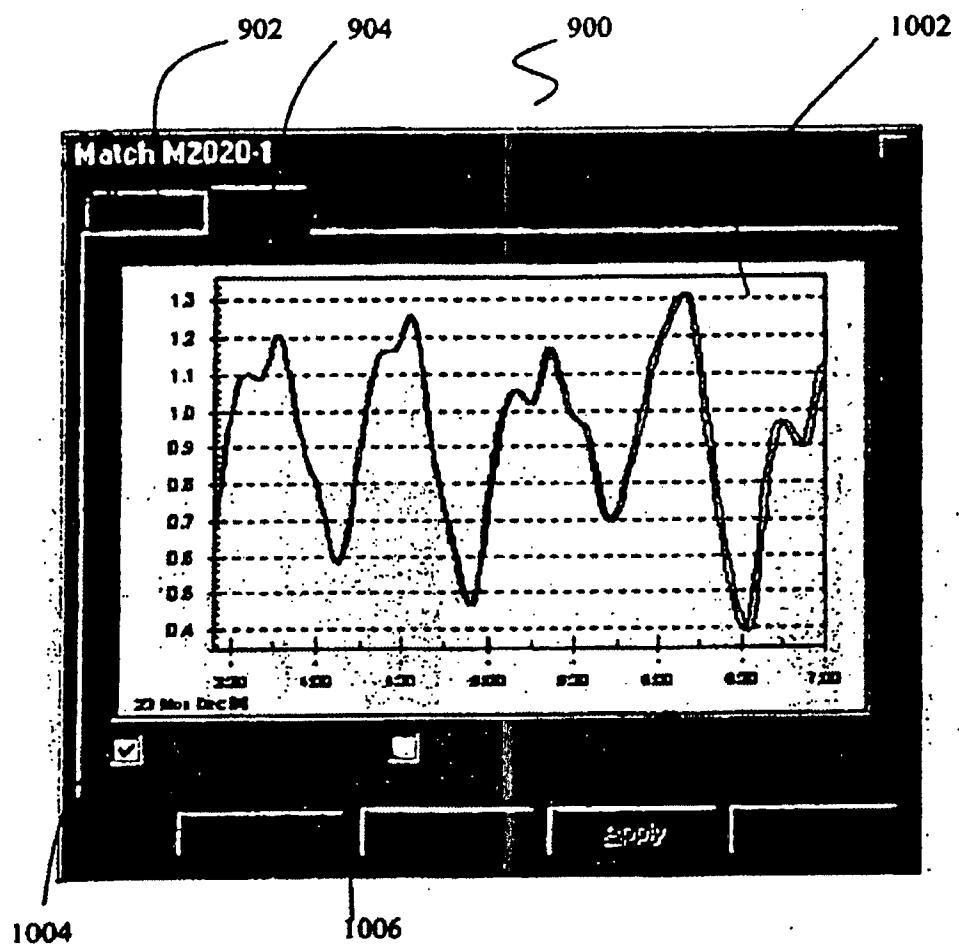


Fig. 10

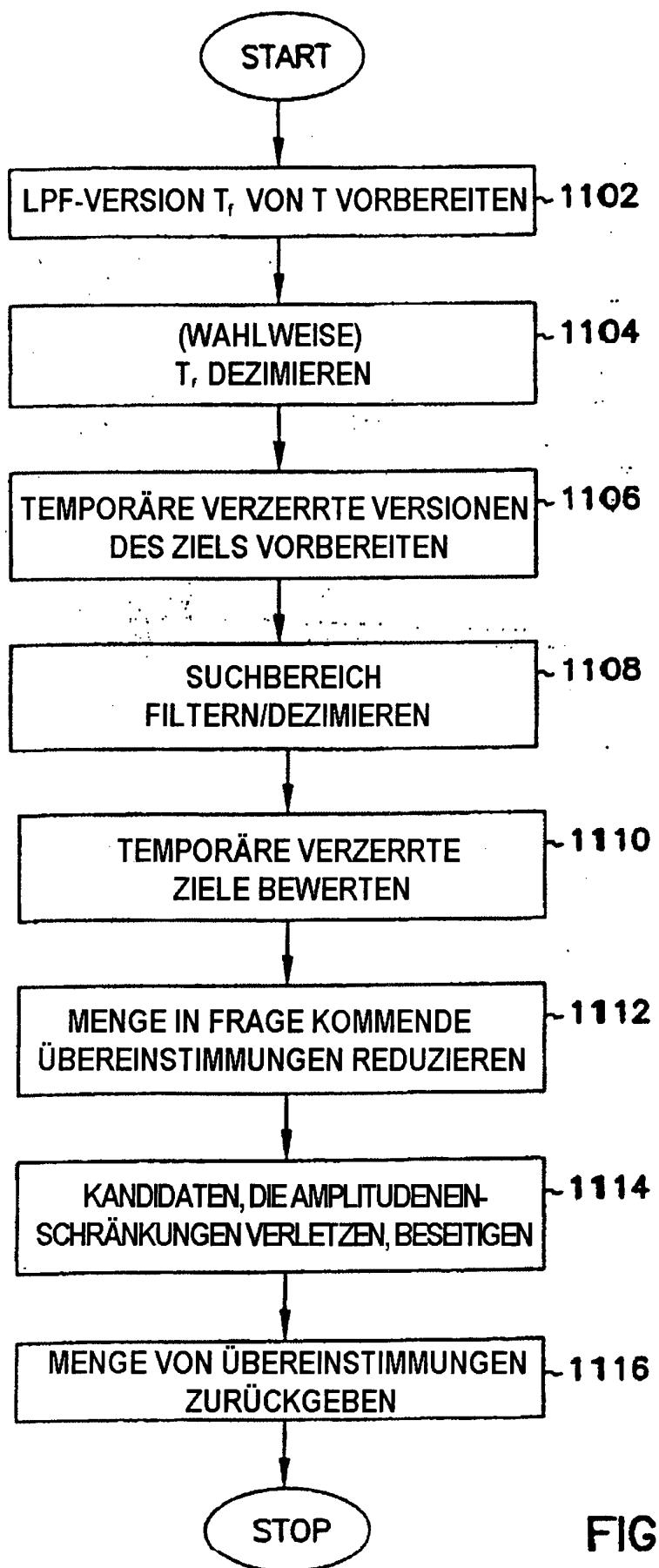


FIG. 11

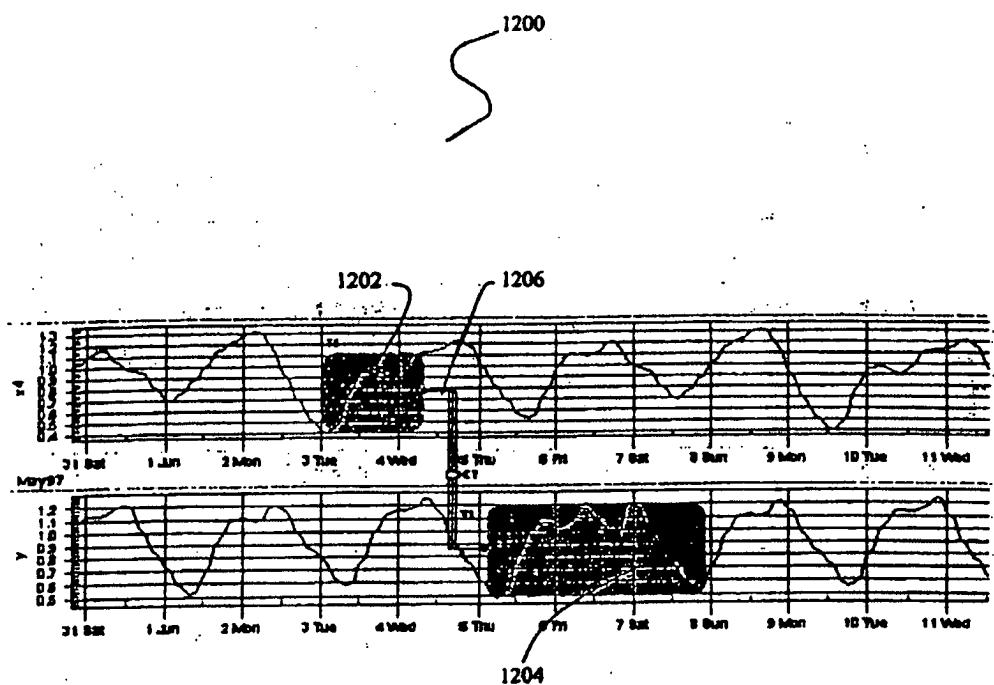


Fig. 12

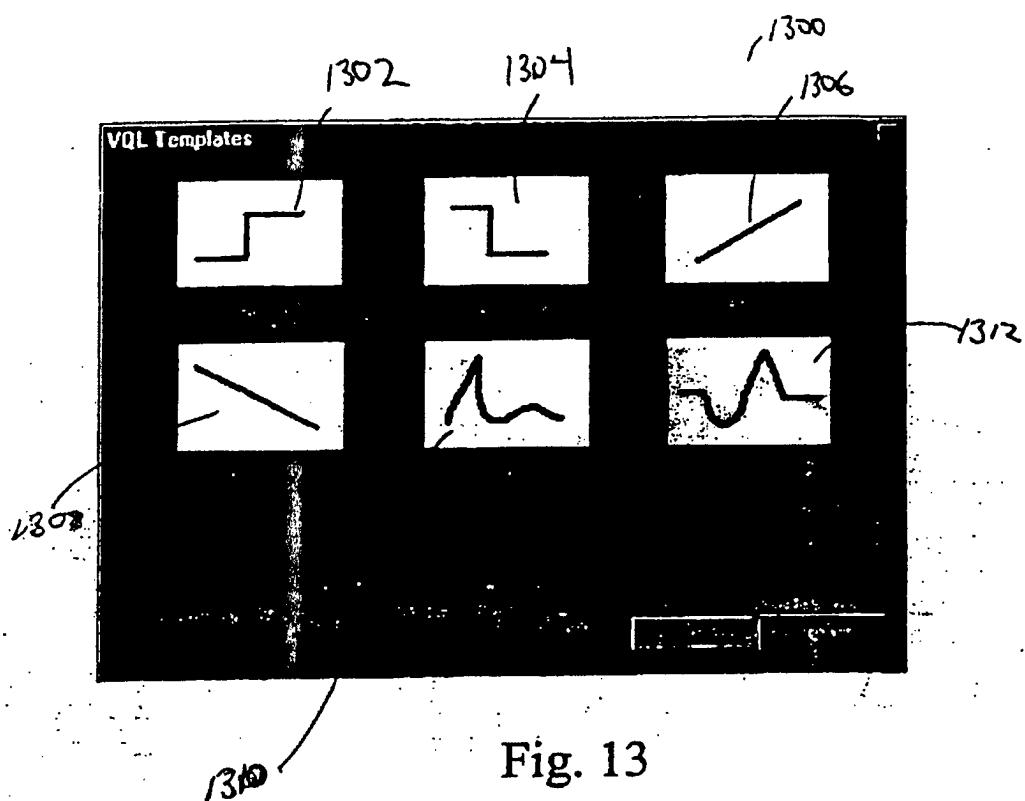


Fig. 13