

Patent De

<http://www.patent-de.com/20060316/DE60020865T2.html>

DE60020865T2

16.03.2006

EP-Veröffentlichungsnummer 0001222656

Titel System, Verfahren und Computerprogramm für einen telefonischen Emotionsdetektor mit Rückmeldung an einen Bediener

Anmelder Accenture LLP, Palo Alto, Calif., US

Erfinder PETRUSHIN, A., Valery, Buffalo Grove IL 60089, US

Vertreter Müller-Boré & Partner, Patentanwälte, European Patent Attorneys, 81671 München

DE-Aktenzeichen 60020865

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

Sprache des Dokument EN

EP-Anmeldetag 31.08.2000

EP-Aktenzeichen 009615469

WO-Anmeldetag 31.08.2000

PCT-Aktenzeichen PCT/US00/24325

WO-Veröffentlichungsnummer 0001016939

WO-Veröffentlichungsdatum 08.03.2001

EP-Offenlegungsdatum 17.07.2002

EP date of grant 15.06.2005

Veröffentlichungstag im Patentblatt 16.03.2006

IPC-Hauptklasse G10L 17/00(2006.01)A, F, I, 20051017, B, H, EP wie es in Anspruch 19 dargelegt ist.

Kurze Beschreibung der Zeichnungen

Die Erfindung wird besser verstanden werden, wenn die folgende detaillierte Beschreibung davon berücksichtigt wird. 30% von ihrem Durchschnitt ist, dann wird von der Zahl bzw.

Nummer des Spitzenfilters Modell und Rahmen zu Rahmen ist. Dieses letzte Merkmal des Rahmenübergangsfehlers erzwingt deshalb Stimmhaftigkeitsübergänge zu den Punkten von maximaler spektraler Änderung. Das Sprachnachrichtensystem beinhaltet ein Spracheingabesignal, welches als eine Zeitserie s_i gezeigt ist, welches an einem LPC-Analyseblock zur Verfügung gestellt wird. Die LPC-Analyse kann durch eine breite Vielzahl konventioneller Techniken vorgenommen werden, jedoch ist das Endprodukt ein Satz von LPC-Parametern und einem Restsignal u_i . Der Hintergrund einer LPC-Analyse im allgemeinen und verschiedener Verfahren zur Extraktion von LPC-Parametern wird in zahlreichen allgemein bekannten Literaturstellen gefunden, beinhaltend Markel und Gray, Linear Prediction of Speech (1976) und Rabiner und Schafer, Digital Processing of Speech Signals (1978) und darin zitierten Literaturstellen. In der gegenwärtig bevorzugten Anordnung wird die analoge Sprachwellenform bei einer Frequenz von 8 KHz und mit einer Genauigkeit von 16 Bits abgetastet, um die Eingabezeitserie s_i zu erzeugen. Natürlich ist das System überhaupt nicht von der Abtastrate oder der verwendeten Genauigkeit abhängig und ist an Sprache, welche bei einer beliebigen Rate abgetastet wurde, oder mit jedem beliebigen Grad von Genauigkeit anwendbar. In der gegenwärtig bevorzugten Anordnung beinhaltet der Satz von LPC-Parametern, welcher verwendet wird, eine Vielzahl von Reflexionskoeffizienten k_i , und ein LPC-Modell der 10. Ordnung wird verwendet (d.h., nur die Reflexionskoeffizienten k_1 bis k_{10} werden extrahiert bzw. entnommen, und Koeffizienten höherer Ordnung werden nicht extrahiert). Jedoch können andere Modellreihenfolgen bzw. -ordnungen oder andere gleichwertige Sätze von LPC-Parametern verwendet werden, wie es für jene mit Erfahrung in der Technik bekannt ist.

Beispielsweise können die LPC-Vorhersagekoeffizienten a_k verwendet werden oder die Impulsantwortabschätzungen e_k . Jedoch sind die Reflexionskoeffizienten k_i am bequemsten. In der gegenwärtig bevorzugten Anordnung werden die Reflexionskoeffizienten gemäß dem Leroux-Gueguen-Verfahren extrahiert, welches beispielsweise in den IEEE Transactions on Acoustic, Speech and Signal Processing, Seite 257 (Juni 1977), dargelegt ist. Jedoch könnten andere Algorithmen, welche jenen mit Erfahrung in der Technik gut bekannt sind, wie beispielsweise Durbin's verwendet werden, um die Koeffizienten zu berechnen. Ein Nebenprodukt der Berechnung der LPC-Parameter wird typischerweise ein Restsignal u_k sein. Jedoch kann, wenn die Parameter durch ein Verfahren berechnet werden, welches nicht automatisch u_k als ein Nebenprodukt ausgibt bzw. hervorbringt, der Rest einfach durch Verwendung der LPC-Parameter gefunden werden, um ein digitales Filter mit endlicher Impulsantwort zu konfigurieren, welches direkt die Restserie u_k aus der eingegebenen bzw. Eingangsserie s_k berechnet. Die Restsignalzeitserie u_k wird nun durch einen sehr einfachen digitalen Filtervorgang gegeben, welcher von den LPC-Parametern für den aktuellen Rahmen abhängig ist. D.h., das Spracheingabesignal s_k ist eine Zeitserie, welche einen Wert aufweist, welcher sich einmal in jeder Abtastung bei einer Abtastrate von z.B. 8 KHz ändern kann. Jedoch werden die LPC-Parameter normalerweise nur einmal in jeder Rahmenperiode bei einer Rahmenfrequenz von z.B. 100 Hz neu berechnet. Das Restsignal u_k ebenfalls weist eine Periode gleich der Abtastperiode auf. Somit ist bzw. wird das Digitalfilter, dessen Wert von den LPC-Parametern abhängig ist, vorzugsweise nicht bei jedem Restsignal u_k neu eingestellt. In der gegenwärtig bevorzugten Anordnung treten ungefähr 80 Werte in der Restsignalzeitserie u_k durch das Filter 14, bevor ein neuer Wert der LPC-Parameter erzeugt wird, und deshalb ist eine neue Charakteristik für das Filter 14 implementiert.

Insbesondere wird der erste Reflexionskoeffizient k_1 aus dem Satz von LPC-Parametern extrahiert, welche durch den LPC-Analyseabschnitt 12 zur Verfügung gestellt sind. Während die LPC-

Parameter selbst die Reflexionskoeffizienten k_1 sind, ist es lediglich notwendig, den ersten Reflexionskoeffizienten k_1 zu suchen bzw. nachzusehen. Jedoch ist, wo andere LPC-Parameter verwendet werden, die Transformation bzw. Umformung der Parameter, um den

Reflexionskoeffizienten erster Ordnung zu erzeugen, typischerweise extrem einfach, beispielsweise, $k_1 = a_1/a_0$. Das System verwendet vorzugsweise den ersten Reflexionskoeffizienten, um 1-poliges adaptives Filter zu definieren. Jedoch muß das Filter nicht ein einpoliges Filter sein, sondern kann als ein komplexeres Filter konfiguriert sein, welches einen oder mehrere Pole oder eine oder mehrere Nullen aufweist, von welchen einige oder alle adaptiv variiert werden können. Es sollte auch beachtet werden, daß die adaptive Filtercharakteristik bzw. Charakteristik des adaptiven Filters nicht durch den ersten Reflexionskoeffizienten k_1 bestimmt werden muß. Wie es in der Technik gut bekannt ist, gibt es zahlreiche gleichwertige Sätze von LPC-Parametern, und die Parameter in anderen LPC-Parametersätzen können ebenfalls wünschenswerte Filtercharakteristika zur Verfügung stellen.

Besonders in einem beliebigen Satz von LPC-Parametern sind die Parameter der niedrigsten Ordnung am wahrscheinlichsten, Information über die grobe spektrale Form zur Verfügung zu stellen. Somit könnte ein adaptives Filter a_1 oder e_1 verwenden, um einen Pol zu definieren, wobei dies ein einziger oder Mehrfachpol sein kann und alleine oder in Kombination mit anderen Nullen und/oder Polen verwendet werden kann. Darüber hinaus muß der Pol (oder null), welcher adaptiv durch ein LPC-Parameter definiert ist, nicht exakt mit diesem Parameter zusammenfallen, sondern kann in Größe oder Phase verschoben sein bzw. werden. Somit filtert das 1-polige adaptive Filter die Restsignalzeitserie u_k , um eine gefilterte Zeitserie u'_k zu erzeugen. Wie oben besprochen, wird diese gefilterte Zeitserie u'_k , ihre hochfrequente Energie bedeutend während der stimmhaften Sprachsegmente reduziert aufweisen, wird jedoch nahezu die vollständige Frequenzbandbreite während der stimmlosen Sprachsegmente beibehalten. Dieses gefilterte Restsignal u'_k wird dann einer weiteren Verarbeitung unterzogen, um die Tonhöhenkandidaten und die Stimmhaftigkeitsentscheidung zu extrahieren.

Eine breite Vielzahl von Verfahren zum Extrahieren der Tonhöheninformation aus einem Restsignal besteht und jede von diesen kann verwendet werden. Viele von diesen werden im allgemeinen in dem oben erwähnten Buch von Markel und Gray besprochen. In der gegenwärtig bevorzugten Anordnung werden die Kandidatentonhöhenwerte erhalten, indem die Spitzen in der normalisierten Korrelationsfunktion des gefilterten Restsignals gefunden wird, definiert wie folgt: für $k_{min}, n \leq k \leq k_{max}$ wo u'_j das gefilterte Restsignal ist, k_{min} und k_{max} die Grenzen für die Korrelationsverzögerung k definieren, und m die Anzahl von Abtastungen in einer Rahmenperiode (80 in der bevorzugten Anordnung) ist und deshalb die Anzahl von zu korrelierenden Abtastungen definiert. Die Kandidatentonhöhenwerte sind bzw. werden durch die Verzögerungen k^* definiert, bei welchem der Wert von $C(k^*)$ ein örtliches Maximum annimmt, und der skalare Wert von $C(k)$ verwendet wird, um einen "Güte"-Wert für jeden Kandidaten k^* zu definieren. Optional wird ein Schwellenwert-Wert C_{min} auf dem Gütemaß $C(k)$ eingeführt, und lokale bzw. örtliche Maxima von $C(k)$, welche nicht den Schwellwert C_{min} überschreiten, werden ignoriert. Wenn kein k^* existiert, für welches $C(k^*)$ größer als C_{min} ist, dann ist der Rahmen notwendigerweise stimmlos.

Alternativ kann der Güteschwellwert C_{min} entfallen und die normalisierte Autokorrelationsfunktion 1112 kann einfach geregelt bzw. gesteuert werden, um eine gegebene Anzahl von Kandidaten auszuweisen, welche die besten Gütewerte aufweisen, z.B. die 16 Tonhöhenperiodenkandidaten k , welche die größten Werte von $C(k)$ aufweisen. In einer Anordnung ist bzw. wird überhaupt kein Schwellwert auf dem Gütewert $C(k)$ eingeführt bzw. diesem überlagert, und keine Stimmhaftigkeitsentscheidung wird auf dieser Stufe vorgenommen. Stattdessen werden die 16 Tonhöhenperiodenkandidaten k^*1, k^*2 usw. zusammen mit dem entsprechenden Gütewert ($C(k^*i)$) für jeden Einzelnen ausgewiesen. In der gegenwärtig bevorzugten Anordnung wird die

Stimmhaftigkeitsentscheidung auf dieser Stufe nicht vorgenommen, selbst wenn alle der $C(k)$ -Werte extrem niedrig sind, jedoch wird die Stimmhaftigkeitsentscheidung vorgenommen in dem nachfolgenden dynamischen Programmierschritt, welcher unten besprochen bzw. diskutiert wird. In der gegenwärtig bevorzugten Anordnung ist bzw. wird eine veränderliche Anzahl von Tonhöhenkandidaten entsprechend einem Spitzenfindungsalgorithmus identifiziert. D.h., der Graph der "Güte"-Werte $C(k)$, verglichen mit der Kandidatentonhöhenperiode k wird nachgeführt bzw. verfolgt.

Jedes örtliche Maximum ist als eine mögliche Spitze identifiziert. Jedoch ist bzw. wird das Vorhandensein einer Spitze an diesem identifizierten lokalen bzw. örtlichen Maximum nicht bestätigt, bis die Funktion danach um einen konstanten Betrag abgefallen ist. Dieses bestätigte lokale Maximum stellt dann einen der Tonhöhenperiodenkandidaten zur Verfügung. Nachdem jeder Spitzenkandidat auf diese Weise identifiziert wurde, sucht der Algorithmus danach nach einem Tal. D.h., jedes lokale Minimum ist bzw. wird als ein mögliches Tal identifiziert, ist jedoch nicht als ein Tal bestätigt, bis die Funktion danach um einen vorbestimmten konstanten Wert angestiegen ist. Die Täler werden nicht getrennt ausgezeichnet bzw. berichtet, jedoch wird ein bestätigtes Tal nötig, nachdem eine bestätigte Spitze vor einer neuen Spitze identifiziert werden wird. In der gegenwärtig bevorzugten Ausführungsform, wo die Gütewerte definiert sind, um durch +1 oder -1 begrenzt zu sein, wurde der zur Bestätigung einer Spitze oder eines Tals erforderliche konstante Wert auf 0,2 eingestellt, jedoch kann dies weitgehend geändert werden. Somit stellt diese Stufe eine variable Anzahl von Tonhöhenkandidaten als Ausgabe, von null bis 15 zur Verfügung. In der gegenwärtig bevorzugten Anordnung wird der Satz von Tonhöhenperiodenkandidaten, welche durch die vorangegangenen Schritte zur Verfügung gestellt werden, dann zu einem dynamischen Programmieralgorithmus zur Verfügung gestellt.

Dieser dynamische Programmieralgorithmus verfolgt dann sowohl Tonhöhen- wie auch Stimmhaftigkeitsentscheidungen, um eine Tonhöhen- und Stimmhaftigkeitsentscheidung für jeden Rahmen bereitzustellen, welcher optimal im Zusammenhang seiner Nachbarn ist. Angesichts der Kandidatentonhöhenwerte und ihrer Gütewerte $C(k)$ wird nun dynamisches Programmieren verwendet, um eine optimale Tonhöhenkontur bzw. -umrißlinie zu erhalten, welche eine optimale Stimmhaftigkeitsentscheidung für jeden Rahmen beinhaltet. Das dynamische Programmieren erfordert, daß verschiedene Sprachrahmen in einem Segment von Sprache analysiert werden, bevor die Tonhöhe und Stimmhaftigkeit für den ersten Rahmen des Segments entschieden werden kann. Bei jedem Rahmen des Sprachsegments wird jeder Tonhöhenkandidat mit den zurückgehaltenen Tonhöhenkandidaten des vorhergehenden Rahmens verglichen.

Jeder zurückgehaltene Tonhöhenkandidat von dem vorhergehenden Rahmen trägt mit sich einen kumulativen bzw. zunehmenden Nachteil bzw. Abzug, und jeder Vergleich zwischen einem neuen Tonhöhenkandidaten und jedem der zurückgehaltenen Tonhöhenkandidaten weist auch ein neues Distanz- bzw. Abstandsmaß auf. Somit gibt es für jeden Tonhöhenkandidaten in dem neuen Rahmen einen kleinsten Nachteil bzw. Abzug, welcher eine beste Übereinstimmung mit einem der bei- bzw. zurückgehaltenen Tonhöhenkandidaten des vorhergehenden Rahmens repräsentiert. Wenn der kleinste kumulative Nachteil für jeden neuen Kandidaten berechnet wurde, wird der Kandidat zusammen mit seinem kumulativen Nachteil und einem Rückwärtszeiger zu der besten Übereinstimmung in dem vorhergehenden Rahmen zurückgehalten. Somit definieren die Rückwärtszeiger eine Trajektorie bzw. Zustandskurve, welche einen kumulativen Abzug aufweist, wie in dem kumulativen Abzugswert des letzten Rahmens in der Projektrate aufgezeichnet wurde. Die optimale Trajektorie für jeden gegebenen Rahmen wird durch ein Auswählen der Trajektorie mit dem minimalen kumulativen Nachteil erhalten.

Der stimmlose Zustand ist als ein Tonhöhenkandidat auf jedem Rahmen definiert. Die Abzugs- bzw. Nachteilsfunktion beinhaltet vorzugsweise Stimmhaftigkeitsinformation, so daß die

Stimmhaftigkeitsentscheidung eine natürliche Folge der dynamischen Programmierstrategie ist. In der gegenwärtig bevorzugten Anordnung ist die dynamische Programmierstrategie 16 breit und 6 tief. D.h., 15 Kandidaten (oder weniger) plus der "Stimmlosigkeits"-Entscheidung (zur Bequemlichkeit als eine Null-Tonhöhenperiode festgelegt) werden als mögliche Tonhöhenperiode auf jedem Rahmen identifiziert, und alle 16 Kandidaten zusammen mit ihren Gütewerten werden für die 6 vorhergehenden Rahmen zurückgehalten.

Die Entscheidungen über Tonhöhe und Stimmhaftigkeit werden endgültig nur in bezug auf den ältesten Rahmen vorgenommen, welcher in dem dynamischen Programmieralgorithmus enthalten ist. D.h., die Tonhöhen- und Stimmhaftigkeitsentscheidung würde die Kandidatentonhöhe beim Rahmen FK-5 akzeptieren, dessen gegenwärtige Trajektorien-Kosten minimal waren. D.h. von den 16 (oder weniger) Trajektorien, welche am neuesten bzw. frischesten Rahmen FK enden, identifiziert die Kandidatentonhöhe im Rahmen FK, welche die geringsten kumulativen Trajektorien-Kosten aufweist, die optimale Trajektorie. Diese optimale Trajektorie wird dann zurückverfolgt und verwendet, um die Tonhöhen/Stimmhaftigkeitsentscheidung für Rahmen FK--5 vorzunehmen. Es soll beachtet werden, daß keine endgültige Entscheidung über die Tonhöhenkandidaten in nachfolgenden Rahmen (Fk-4 usw.) vorgenommen wird, da die optimale Trajektorie nicht länger optimal erscheinen mag, nachdem mehrere Rahmen bewertet sind.

Natürlich kann, wie es für jene mit Erfahrung in der Technik und der numerischen Optimierung wohl bekannt ist, eine endgültige Entscheidung in einem derartigen dynamischen Programmieralgorithmus alternativ zu anderen Zeiten bzw. Zeitpunkten vorgenommen werden, z.B. im nächsten bis letzten Rahmen, der in dem Puffer gehalten wird. Zusätzlich können die Breite und Tiefe des Puffers weitgehend variiert werden. Beispielsweise können bis zu 64 Tonhöhenkandidaten bewertet werden oder sowenig wie zwei; der Puffer könnte sowenig wie einen vorigen Rahmen zurückhalten, oder soviel wie 16 vorige Rahmen oder mehr, und andere Modifikationen und Abänderungen können eingerichtet werden, wie durch jene mit Erfahrung in der Technik erkannt werden wird.

Der dynamische Programmieralgorithmus ist bzw. wird durch den Übergangsfehler zwischen einem Tonhöhenperiodenkandidaten in einem Rahmen und einem anderen Tonhöhenperiodenkandidaten in dem nachfolgenden Rahmen definiert. In der gegenwärtig bevorzugten Anordnung ist dieser Übergangsfehler als die Summe von drei Teilen definiert: einem Fehler EP aufgrund von Tonhöhenabweichungen, einem Fehler Es aufgrund von Tonhöhenkandidaten, welche einen niedrigen "Güte"-Wert aufweisen, und einem Fehler E1 aufgrund des Stimmhaftigkeitsübergangs. Der Tonhöhenabweichungsfehler EP ist eine Funktion der gegenwärtigen Tonhöhenperiode und der vorigen Tonhöhenperiode, angegeben durch: wenn beide Rahmen stimmhaft sind, und $EP = BP \cdot \tau \cdot DN$ andernfalls; wo τ die Kandidatentonhöhenperiode des gegenwärtigen Rahmens ist, τ eine zurückgehaltene Tonhöhenperiode des vorigen Rahmens, in bezug auf welchen der Übergangsfehler berechnet wird, ist, und BP, AD und DN Konstante sind. Es soll beobachtet werden, daß die Minimumfunktion eine Vorkehrung zur Tonhöhenperiodenverdopplung und Tonhöhenperiodenhalbierung beinhaltet. Diese Vorkehrung ist nicht unbedingt notwendig, wird jedoch als vorteilhaft betrachtet. Natürlich könnte optional eine ähnliche Vorkehrung zur Tonhöhenperiodenverdreifachung beinhaltet sein, usw. Der Stimmhaftigkeitszustandsfehler ES ist eine Funktion des "Güte"-Werts $C(k)$ des gegenwärtig betrachteten Rahmentonhöhenkandidaten.

Für den stimmlosen Kandidaten, welcher immer unter den 16 oder weniger Tonhöhenperiodenkandidaten enthalten ist, die für jeden Rahmen zu berücksichtigen sind, wird der Gütewert $C(k)$ gleich dem Maximum von $C(k)$ für alle anderen 15 Tonhöhenperiodenkandidaten im selben Rahmen gesetzt. Der Stimmhaftigkeitszustandsfehler ES ist gegeben durch $ES = BS(R_v - C(\tau))$, wenn der gegenwärtige Kandidat stimmhaft ist, und $ES = BS(C(\tau) - RU)$ andernfalls, wo $C(\tau)$ der "Gütewert" entsprechend dem gegenwärtigen Tonhöhenkandidaten τ ist, und BS, R_v ,

und RU Konstante sind. Der Stimmhaftigkeitsübergangsfehler ET ist im Hinblick auf ein spektrales Differenzmaß T definiert. Das spektrale Differenzmaß T definierte, für jeden Rahmen, im allgemeinen, wie unterschiedlich sein Spektrum von dem Spektrum des aufnehmenden Rahmens ist. Offensichtlich könnte eine Anzahl von Definitionen für ein derartiges spektrales Differenzmaß verwendet werden, welches in der gegenwärtig bevorzugten Anordnung definiert ist wie folgt: wo E die RMS-Energie bzw. Effektivwertenergie des gegenwärtigen Rahmens ist, EP die Energie des vorigen Rahmens ist, L(N) ist der N-te Logarithmus des Flächenverhältnisses des augenblicklichen Rahmens und LP(N) N-te Logarithmus des Flächenverhältnisses des vorigen Rahmens ist. Das logarithmische Flächenverhältnis L(N) wird direkt aus dem N-ten Reflexionskoeffizienten kN berechnet wie folgt: Der Stimmhaftigkeitsübergangsfehler ET ist dann als eine Funktion des spektralen Differenzmaßes definiert wie folgt:

Wenn der gegenwärtige und vorige Rahmen beide stimmlos sind, oder wenn beide stimmhaft sind, wird ET gleich 0 gesetzt; andernfalls, $ET = GT + AT/T$, wo T das spektrale Differenzmaß bzw. Maß der spektralen Differenz des gegenwärtigen Rahmens ist. Wieder könnte die Definition des Stimmhaftigkeitsübergangsfehlers weitgehend variiert werden. Das Schlüsselmerkmal des wie hier definierten Stimmhaftigkeitsübergangsfehlers ist, daß, wann immer eine Stimmhaftigkeitszustandsänderung (stimmhaft zu stimmlos oder stimmlos zu stimmhaft) auftritt, ein Nachteil bzw. Abzug festgestellt wird, welcher eine abnehmende Funktion der spektralen Differenz zwischen den zwei Rahmen ist. D.h., eine Änderung im Stimmhaftigkeitszustand wird mißbilligt, außer es tritt auch ein beträchtlicher spektraler Wechsel auf.

Eine derartige Definition eines Stimmhaftigkeitsübergangsfehlers liefert beträchtliche Vorteile, da sie die Verarbeitungszeit reduziert, die erforderlich ist, um ausgezeichnete Stimmhaftigkeitsentscheidungen zu liefern. Die anderen Fehler ES und EP, welche den Übergangsfehler in der gegenwärtig bevorzugten Anordnung ausmachen, können ebenfalls verschiedenartig definiert sein. D.h., der Stimmhaftigkeitszustandsfehler kann auf jede Weise definiert sein, welche im allgemeinen Tonhöhenperiodenhypothesen, welche zu den Daten in dem gegenwärtigen Rahmen zu passen scheinen, gegenüber jenen bevorzugt, welche weniger gut zu den Daten passen. In ähnlicher Weise kann der Tonhöhenabweichungsfehler EP auf jede Weise definiert sein, welche im allgemeinen mit Änderungen in der Tonhöhenperiode übereinstimmt bzw. diesen entspricht. Es ist nicht notwendig für den Tonhöhenabweichungsfehler, eine Vorkehrung zum Verdoppeln und Halbieren zu beinhalten, wie dies hier festgelegt wurde, obwohl eine derartige Vorkehrung wünschenswert ist.

Ein weiteres optionales Merkmal ist, daß, wenn der Tonhöhenabweichungsfehler Vorkehrungen zum Nachfolgen der Tonhöhe über Verdopplungen und Halbierungen beinhaltet, es wünschenswert sein kann, die Tonhöhenperiodenwerte entlang der optimalen Trajektorie zu verdoppeln (oder zu halbieren), nachdem die optimale Trajektorie identifiziert wurde, um diese so weit wie möglich konsistent bzw. einheitlich zu machen. Es sollte beachtet werden, daß es nicht notwendig ist, alle der drei identifizierten Komponenten des Übergangsfehlers zu verwenden.

Beispielsweise könnte der Stimmhaftigkeitszustandsfehler ausgelassen sein, wenn einige vorige Stufen Tonhöhenhypothesen mit einem niedrigen "Güte"-Wert ausschlossen, oder wenn die Tonhöhenperioden auf nach dem "Güte"-Wert in einer gewissen Weise derart geordnet wurden, daß die Tonhöhenperioden, welche einen höheren Gütewert aufweisen, bevorzugt würden, oder auf andere Weise. In ähnlicher Weise können andere Komponenten in der Übergangsfehlerdefinition nach Wunsch beinhaltet sein. Es sollte auch beachtet werden, daß das dynamische Programmierverfahren, welches hierin gelehrt wurde, nicht notwendigerweise auf Tonhöhenperiodenkandidaten angewendet werden muß, welche aus einem adaptiv gefilterten Restsignal extrahiert wurden, noch auf Tonhöhenperiodenkandidaten, welche aus dem LPC-Restsignal abgeleitet wurden, sondern auf jeden Satz von Tonhöhenperiodenkandidaten angewendet

werden kann, welche Tonhöhenperiodenkandidaten beinhalten, die direkt aus dem ursprünglichen Eingabesprachsignal extrahiert wurden.

Diese drei Fehler werden dann summiert, um den Gesamtfehler zwischen irgendeinem Tonhöhenkandidaten in dem gegenwärtigen Rahmen und irgendeinem Tonhöhenkandidaten in dem vorhergehenden Rahmen zur Verfügung zu stellen. Wie oben angemerkt wurde, werden diese Übergangsfehler dann kumulativ summiert, um kumulative Nachteile bzw. Abzüge für jede Trajektorie in dem dynamischen Programmieralgorithmus zur Verfügung zu stellen. Dieses dynamische Programmierverfahren für ein gleichzeitiges Finden von sowohl Tonhöhe wie auch Stimmhaftigkeit ist in sich selbst neuartig, und muß nicht nur in Kombination mit den gegenwärtig bevorzugten Verfahren eines Findens von Tonhöhenperiodenkandidaten verwendet werden. Jedes Verfahren zum Finden von Tonhöhenkandidaten kann in Kombination mit diesem neuartigen dynamischen Programmieralgorithmus verwendet werden. Egal welches Verfahren verwendet wird, um Tonhöhenperiodenkandidaten zu finden, werden die Kandidaten einfach als Eingabe zu dem dynamischen Programmieralgorithmus zur Verfügung gestellt. Insbesondere ist, während ein Verwenden eines Minicomputers und ein hochgenaues Abtasten gegenwärtig bevorzugt ist, dieses System nicht wirtschaftlich für großvolumige Anwendungen. Somit wird von dem bevorzugten System in der Zukunft erwartet, eine Anordnung zu sein, welche ein auf einem Mikrocomputer basierendes System verwendet, wie beispielsweise den TI Professional Computer. Dieser professionelle Computer, wenn er mit einem Mikrofon, Lautsprecher und einer Sprachverarbeitungskarte konfiguriert ist, beinhaltend einen TMS 320 numerisch verarbeitenden Mikroprozessor und Datenwandler, ist ausreichende Hardware, um das System zu realisieren.

Anspruch[de] 1. Verfahren zum Überwachen von Emotionen in Sprech- bzw. Stimmsignalen und Bereitstellen einer Rückmeldung bzw. eines Feedbacks darauf, umfassend die Schritte eines: (a) Empfangens eines Stimmsignals, das für eine Komponente einer Konversation zwischen wenigstens zwei Menschen bzw. Personen repräsentativ ist; (b) Segmentierens des Stimmsignals in Bereiche bzw. Regionen über einen Bereich einer Tonfrequenz des Stimmsignals; (c) Bestimmens einer Emotion, die mit dem Stimmsignal assoziiert ist, durch Verwendung der Bereiche als ein Eingabevektor zu einem neuronalen Netzwerk, das verwendet wird, um die Emotion zu bestimmen; und (d) Bereitstellens einer Rückmeldung zu einer dritten Partei, basierend auf der Emotion, welche aus den Bereichen bestimmt wird, die in das neurale Netzwerk eingegeben wird. 2. Verfahren nach Anspruch 1, wobei die Rückmeldung nur bereitgestellt wird, wenn die Emotion als eine negative Emotion bestimmt wird, die aus der Gruppe von negativen Emotionen, bestehend aus Zorn, Traurigkeit und Angst gewählt ist. 3. Verfahren nach Anspruch 1, wobei die Emotion durch ein Extrahieren eines Merkmals aus dem Stimmsignal bestimmt wird. 4. Verfahren nach einem der Ansprüche 1 bis 3, wobei die Konversation über ein Telekommunikationsnetzwerk ausgeführt wird. 5. Verfahren nach Anspruch 4, wobei eine der Personen ein Kunde, eine der Personen ein Angestellter und die dritte Partei ein Manager ist. 6. Verfahren nach Anspruch 5, wobei der Angestellte durch ein Telefonzentrum bzw. Callcenter angestellt wird. 7. Verfahren nach einem der Ansprüche 1 bis 6, wobei das neurale Netzwerk ein zweilagiges rückwärts fortschreitendes bzw. fortpflanzendes neurales Netzwerk ist. 8. Verfahren nach einem der Ansprüche 1 bis 6, wobei ein Bestimmen einer Emotion ein Treffen von Entscheidungen basierend auf einem Abstimm- bzw. Stimmverhalten mit einer ungeraden Anzahl von Klassifizierern des neuronalen Netzwerks ist. 9. Verfahren nach einem der Ansprüche 1 bis 8, wobei ein Segmentieren des Stimm- bzw. Sprachsignals ein Filtern des Stimmsignals mit einem elektronischen Filter beinhaltet. 10. Computerprogramm, das auf einem computerlesbarem Medium verkörpert ist, zum Überwachen von Emotionen in Sprach- bzw. Stimmsignalen und Bereitstellen einer Rückmeldung bzw. Feedbacks darauf, wenn das Computerprogramm auf einem Computer läuft, wobei das Computerprogramm umfaßt: (a) ein Codesegment, welches ein Stimmsignal erhält bzw. empfängt, das für eine Komponente einer Konversation zwischen wenigstens zwei Menschen bzw. Personen repräsentativ ist;

(b) ein Codesegment, welches das Stimmsignal in Bereiche bzw. Regionen über einen Bereich einer Tonfrequenz des Stimmsignals segmentiert bzw. unterteilt; (c) ein Codesegment, welches eine Emotion bestimmt, die mit dem Stimmsignal assoziiert ist, indem die Bereiche als ein Eingabevektor zu einem neuronalen Netzwerk verwendet werden, das verwendet wird, um die Emotion zu bestimmen; und (d) ein Codesegment, welches eine Rückmeldung bzw. Feedback zu einer dritten Partei basierend auf der Emotion bereitstellt, welche aus den Bereichen bestimmt ist, die in das neurale Netzwerk eingegeben sind. 11. Computerprogramm nach Anspruch 10, wobei die Rückmeldung nur bereitgestellt ist, wenn die Emotion als eine negative Emotion bestimmt ist, gewählt aus der Gruppe von negativen Emotionen, bestehend aus Zorn, Traurigkeit und Angst. 12. Computerprogramm nach Anspruch 10, wobei die Emotion durch ein Extrahieren eines Merkmals aus dem Stimmsignal bestimmt ist. 13. Computerprogramm nach einem der Ansprüche 10 bis 12, wobei die Konversation über ein Telekommunikationsnetzwerk ausgeführt wird. 14. Computerprogramm nach Anspruch 13, wobei eine der Personen ein Kunde, eine Person ein Angestellter ist, und die dritte Partei ein Manager ist. 15. Computerprogramm nach Anspruch 14, wobei der Angestellte durch ein Callcenter bzw. ein Telefonzentrum angestellt ist. 16. Computerprogramm nach einem der Ansprüche 10 bis 15, wobei das neurale Netzwerk ein zweilagiges rückwärts fortschreitendes bzw. fortpflanzendes neuronales Netzwerk ist. 17. Computerprogramm nach einem der Ansprüche 10 bis 15, wobei eine Emotion durch das Codesegment bestimmt ist, basierend auf einem Abstimm- bzw. Stimmverhalten mit einer ungeraden Anzahl von Klassifizierern des neuronalen Netzwerks bestimmt ist. 18. Computerprogramm nach einem der Ansprüche 10 bis 17, wobei das Stimmsignal durch ein Codesegment segmentiert ist, welches eine Filtration des Stimmsignals mit einem elektronischen Filter anleitet. 19. System zum Überwachen von Emotionen in Stimm- bzw. Sprachsignalen und Bereitstellen einer Rückmeldung bzw. Feedback dafür, umfassend: (a) Logikmittel, die adaptiert sind, um das Stimmsignal zu empfangen, das für eine Komponente einer Konversation zwischen wenigstens zwei Menschen bzw. Personen repräsentativ ist; (b) Logikmittel, die adaptiert sind, um das Stimmsignal in Bereiche bzw. Regionen über einen Bereich einer Tonfrequenz des Stimmsignals zu segmentieren; (c) Logikmittel, die adaptiert sind, um eine Emotion zu bestimmen, die mit dem Stimmsignal assoziiert bzw. diesem zugeordnet ist, indem die Regionen als ein Eingabevektor zu einem neuronalen Netzwerk verwendet werden, welches zur Bestimmung der Emotion verwendet ist, und (d) Logikmittel, die adaptiert sind, um eine Rückmeldung zu einer dritten Partei basierend auf der Emotion bereitzustellen, welche aus den Bereichen bestimmt ist, die in das neurale Netzwerk eingegeben sind. 20. System nach Anspruch 19, wobei die Rückmeldung nur bereitgestellt wird, wenn die Emotion als eine negative Emotion bestimmt ist, ausgewählt aus der Gruppe von negativen Emotionen, bestehend aus Zorn, Traurigkeit und Angst. 21. System nach Anspruch 19, wobei die Emotion durch ein Extrahieren eines Merkmals aus dem Stimmsignal bestimmt ist. 22. System nach einem der Ansprüche 19 bis 21, wobei die Konversation über ein Telekommunikationsnetzwerk ausgeführt ist. 23. System nach Anspruch 22, wobei eine der Personen ein Kunde ist, eine der Personen ein Angestellter ist, und die dritte Partei ein Manager ist. 24. System nach Anspruch 23, wobei der Angestellte durch ein Telefonzentrum bzw. Callcenter angestellt ist. 25. System nach einem der Ansprüche 19 bis 24, wobei das neurale Netzwerk ein zweilagiges rückwärts fortschreitendes bzw. fortpflanzendes neuronales Netzwerk ist. 26. System nach einem der Ansprüche 19 bis 24, wobei eine Emotion durch eine Logik bestimmt ist, basierend auf einem Stimmverhalten mit einer ungeraden Anzahl von Klassifizierern des neuronalen Netzwerks. 27. System nach einem der Ansprüche 19 bis 26, wobei das Stimmsignal durch eine Logik segmentiert ist, welche eine Filtration des Stimmsignals mit einem elektronischen Filter anleitet. Es folgen 17 Blatt Zeichnungen

Gebiet der Erfindung

Die vorliegende Erfindung bezieht sich auf Spracherkennung und insbesondere auf das Bereitstellen einer Rückkopplung, basierend auf erkennbaren Stimmcharakteristika bzw. -merkmalen, welche während einer Konversation extrahiert werden.

Hintergrund der Erfindung

Obwohl die erste Monographie über ein Ausdrücken von Emotionen bei Tieren und Menschen durch Charles Darwin im letzten Jahrhundert geschrieben wurde und Psychologen allmählich Kenntnis auf dem Gebiet der Emotionsdetektion und der Stimmerkennung gesammelt haben, hat es eine neue Welle von Interesse kürzlich sowohl von Psychologen wie auch Spezialisten für künstliche Intelligenz auf sich gezogen. Es gibt verschiedene Gründe für dieses erneuerte Interesse: technologischen Fortschritt beim Aufzeichnen, Speichern und Verarbeiten von audiovisueller Information; die Entwicklung von nicht-eindringenden Sensoren; die Einführung bzw. das Auftreten von tragbaren Computern; und den Zwang, die Mensch-Computer-Schnittstelle vom Zeigen und Klicken zu einem Empfinden und Fühlen zu erweitern. Weiters ist es als ein neues Gebiet von Forschung, die in der AI bzw. künstlichen Intelligenz als gemütsbezogenes bzw. emotionales Berechnen bekannt ist, kürzlich identifiziert worden.

Betreffend Forschung beim Erkennen von Emotionen in der Sprache haben einerseits Psychologen viele Experimente durchgeführt und Theorien vorgeschlagen. Andererseits trugen AI-Forscher zu den folgenden Gebieten bei: Synthese von emotionaler Sprache, Erkennung von Emotionen und die Verwendung von Agentien bzw. Mitteln zum Dekodieren und Ausdrücken von Emotionen. Ein ähnlicher Fortschritt wurde bei der Spracherkennung gemacht.

Trotz der Erforschung beim Erkennen von Emotionen in der Sprache war die Technik ohne Verfahren und Vorrichtungen, welche eine Emotionserkennung und Stimm- bzw. Spracherkennung für Geschäftszwecke verwenden.

WO-A-99/31653 beschreibt eine Vorrichtung zum Detektieren des emotionalen Zustands eines Individuums unter Verwendung eines Stimmanalysators, um eine Sprachprobe einzugeben, welche durch das Individuum erzeugt wurde, und um eine Internations- bzw. Sprachmelodieinformation davon abzuleiten. WO-A-99/22364 leitet einen Emotionsgehalt aus dem Audiosignal in der Frequenzdomäne ab und schlägt die Verwendung eines neuronalen Netzwerks zur Emotionsklassifizierung vor.

Gemäß der vorliegenden Erfindung werden ein Verfahren, wie es in Anspruch 1 dargelegt ist, ein Computerprogramm, wie es in Anspruch 10 dargelegt ist, und ein System zur Verfügung gestellt,

Eine solche Beschreibung nimmt bezug auf die beigelegten Zeichnungen, worin:

1 ein schematisches Diagramm einer Hardware-Implementierung von einer Ausführungsform der vorliegenden Erfindung ist;

2 ein Flußdiagramm ist, welches eine Ausführungsform der vorliegenden Erfindung darstellt, welche Emotion unter Verwendung einer Stimm- bzw. Sprachanalyse detektiert;

3 ein Graph ist, welcher die durchschnittliche bzw. Durchschnittsgenauigkeit einer Erkennung für

einen s70-Datensatz zeigt;

4 eine Karte ist, welche die Durchschnittsgenauigkeit einer Erkennung für einen s80-Datensatz illustriert;

5 ein Graph ist, welcher die Durchschnittsgenauigkeit einer Erkennung für einen s90-Datensatz darstellt;

6 ein Flußdiagramm ist, welches eine Ausführungsform der vorliegenden Erfindung illustriert, welche eine Emotion unter Verwendung von Statistik detektiert;

7 ein Flußdiagramm ist, welches ein Verfahren zum Detektieren von Nervosität in einer Stimme in einer Geschäftsumgebung illustriert, um zu helfen, einen Betrug zu verhindern;

8 ein Flußdiagramm ist, welches eine Vorrichtung zum Detektieren von Emotion aus einer Stimmprobe in Übereinstimmung mit einer Ausführungsform der vorliegenden Erfindung darstellt;

9 ein Flußdiagramm ist, welches eine Vorrichtung zum Erzeugen sichtbarer Aufzeichnungen von Schall in Übereinstimmung mit einer Ausführungsform der Erfindung illustriert;

10 ein Flußdiagramm ist, welches eine Ausführungsform der vorliegenden Erfindung illustriert, welche Emotionen in Stimmsignalen überwacht und eine Rückkopplung basierend auf den detektierten Emotionen zur Verfügung stellt;

11 ein Flußdiagramm ist, welches ein System illustriert, welches Anwender- gegen Computeremotionsdetektion von Stimmsignalen vergleicht, um eine Emotionserkennung von entweder einer Ausführungsform der Erfindung, einem Anwender, oder beiden zu verbessern;

12 ein schematisches Diagramm in Blockform einer Spracherkennungsvorrichtung ist;

13 ein schematisches Diagramm in Blockform der Elementanordnung und eines Speicherblocks in 12 ist;

14 ein Spracherkennungssystem mit einem Biomonitor und einem Vorprozessor illustriert;

15 ein Biosignal illustriert, welches durch den Biomonitor von 14 erzeugt wurde;

16 einen Schaltkreis innerhalb des Biomonitors illustriert;

17 ein Blockdiagramm des Vorprozessors ist;

18 eine Beziehung zwischen einer Tonhöhenmodifizierung und dem Biosignal illustriert;

19 ein Flußdiagramm eines Kalibrationsprogramms ist;

20 allgemein die Konfiguration des Abschnitts des Systems zeigt, wobei eine verbesserte Auswahl eines Satzes von Tonhöhenperiodenkandidaten erhalten bzw. erzielt wird.

Detaillierte Beschreibung

In Übereinstimmung mit wenigstens einer Ausführungsform der vorliegenden Erfindung ist bzw. wird ein System zum Durchführen verschiedener Funktionen und Aktivitäten durch Stimmanalyse

und Stimmerkennung zur Verfügung gestellt. Das System kann eingerichtet sein, indem eine Hardware-Implementierung, wie beispielsweise jene verwendet wird, die in 1 illustriert ist. Weiters können verschiedene funktionelle und Anwender-Schnittstellenmerkmale einer Ausführungsform der vorliegenden Erfindung eingerichtet bzw. ermöglicht sein, indem eine Softwareprogrammierung, z.B. objektorientierte Programmierung (OOP) verwendet wird.

Hardwareüberblick

Eine repräsentative Hardwareumgebung einer bevorzugten Ausführungsform der vorliegenden Erfindung ist in 1 dargestellt, welche eine typische Hardwarekonfiguration einer Arbeitsstation illustriert, welche eine Zentralrecheneinheit 110, wie beispielsweise einen Mikroprozessor, und eine Anzahl von anderen Einheiten aufweist, welche über einen Systembus 112 verbunden sind. Die in 1 gezeigte Arbeitsstation beinhaltet einen Direktzugriffsspeicher bzw. Schreib-Lese-Speicher (RAM) 114, einen Nurlese- bzw. Festwertspeicher (ROM) 116, einen I/O-Adapter 118 zum Verbinden bzw. Anschließen peripherer Geräte, wie beispielsweise Diskettenspeichereinheiten 120 mit dem Bus 112, einen Anwender-Schnittstellenadapter 122 zum Verbinden einer Tastatur 124, einer Maus 126, eines Lautsprechers 128, eines Mikrophons 132, und/oder andere Anwender-Schnittstellengeräte, wie beispielsweise einen Berührungsbildschirm bzw. Sensorbildschirm (nicht gezeigt) mit dem Bus 112, Kommunikationsadapter 134 zum Verbinden der Arbeitsstation mit einem Kommunikationsnetzwerk (z.B. einem Datenverarbeitungsnetzwerk) und einen Anzeigeadapter 136 zum Verbinden des Buses 112 mit einer Anzeigevorrichtung 138. Die Arbeitsstation weist typischerweise darauf angesiedelt ein Betriebssystem auf, wie beispielsweise das Microsoft Windows NT oder Windows/95 Betriebssystem (OS), das IBM OS/2-Betriebssystem, das MAC OS, oder UNIX-Betriebssystem.

Emotionserkennung

Die vorliegende Erfindung ist auf ein Verwenden einer Erkennung von Emotionen in der Sprache für Geschäftszwecke gerichtet. Einige Ausführungsformen der vorliegenden Erfindung können verwendet werden, um die Emotion einer Person basierend auf einer Stimmanalyse zu detektieren und die detektierte Emotion der Person auszugeben. Andere Ausführungsformen der vorliegenden Erfindung können für die Detektion des emotionalen Zustands in Telefon-Call-Center-Unterhaltungen und ein Bereitstellen einer Rückkopplung bzw. eines Feedbacks für einen Betreiber oder Überwacher für Überwachungszwecke verwendet werden.

Wenn die Zielsubjekte bekannt sind, wird vorgeschlagen, daß eine Studie an einigen der Zielsubjekte ausgeführt wird, um zu bestimmen, welche Abschnitte einer Stimme am verlässlichsten als Indikatoren bzw. Anzeiger einer Emotion sind. Wenn Zielsubjekte nicht verfügbar sind, können andere Subjekte bzw. Personen verwendet werden. Unter Berücksichtigung dieser Orientierung gilt für die folgende Diskussion:

- * • Daten sollten von Leuten gefordert bzw. gesammelt werden, welche nicht professionelle Schauspieler oder Schauspielerinnen sind, um die Genauigkeit zu verbessern, da Schauspieler und Schauspielerinnen eine bestimmte Sprachkomponente überbetonen könnten, was einen Fehler erzeugt.

- * • Daten könnten von Testsubjekten gefordert werden, welche aus einer Gruppe ausgewählt sind, von welcher erwartet wird, daß sie analysiert wird. Dies würde die Genauigkeit verbessern.

- * • Auf Sprache in Telefonqualität (< 3,4 kHz) kann abgezielt werden, um eine Genauigkeit zur Verwendung mit einem Telefonsystem zu verbessern. modernen Spracherkennungstechniken ausgeschlossen würden, da diese eine viel bessere Qualität des Signals und Rechenleistung

erfordern.

Datensammlung & Evaluierung

In einem beispielhaften Test werden vier kurze Sätze von jedem von dreißig Leuten aufgezeichnet bzw. aufgenommen:

- * • "Dies ist nicht, was ich erwartete."
- * • "Ich werde da sein."
- * • "Morgen ist mein Geburtstag."
- * • "Ich werde nächste Woche heiraten."

Jeder Satz sollte fünf mal aufgezeichnet werden; jedesmal porträtiert das Subjekt einen der folgenden emotionalen Zustände: Fröhlichkeit, Ärger, Traurigkeit, Angst/Nervosität und normal (unemotionell). Fünf Subjekte können auch die Sätze zweimal mit unterschiedlichen Aufzeichnungsparametern aufzeichnen. Somit hat jedes Subjekt 20 oder 40 Aussagen aufgezeichnet, welche einen Bestand ergeben, der 700 Aussagen mit 140 Aussagen pro emotionalem Zustand enthält. Jede Aussage kann unter Verwendung eines Nahebesprechungsmikrofons aufgezeichnet werden; die ersten 100 Aussagen bei 22-kHz/8 Bit und die verbleibenden 600 Aussagen bei 22-kHz/16 Bit.

Nach Erstellen des Bestands kann ein Experiment ausgeführt werden, um die Antworten auf die folgenden Fragen zu finden:

- * • Wie gut können Leute ohne spezielles Training Emotionen in der Sprache porträtieren bzw. darstellen und erkennen?
- * • Wie gut können Leute ihre eigene Emotionen erkennen, welche sie 6–8 Wochen früher aufzeichneten?
- * • Welche Arten von Emotionen sind leichter/schwerer zu erkennen?

Ein wichtiges Ergebnis des Experiments ist eine Auswahl eines Satzes der zuverlässigsten Aussagen, d.h. Aussagen bzw. Äußerungen, die durch die meisten Leute erkannt werden. Dieser Satz kann als Trainings- und Testdaten für Musterbekenennungsalgorithmen verwendet werden, die auf einem Computer laufen.

Ein interaktives Programm eines Typs, welches in der Technik bekannt ist, kann verwendet werden, um die Aussagen in zufälliger Reihenfolge auszuwählen und wiederzugeben und es einem Anwender zu gestatten, jede Aussage entsprechend ihrem emotionalen Inhalt zu klassifizieren. Beispielsweise können dreiundzwanzig Subjekte bzw. Personen an der Evaluierungsstufe und zusätzliche 20 von jenen teilnehmen, welche früher im Aufnahmezustand teilgenommen haben.

Tabelle 1 zeigt eine Leistungs-Verwirrungsmatrix, welche aus Daten resultiert, die aus der Darbietung der zuvor besprochenen Studie gesammelt hat. Die Reihen und Spalten repräsentieren jeweils wahre und bewertete Kategorien. Beispielsweise sagt die zweite Reihe aus, daß 11,9% von Aussagen, welche als glücklich porträtiert wurden, als normal (unemotional) bewertet wurden, 61,4% als wirklich glücklich, 10,1% als ärgerlich, 4,1% als traurig, und 12,5% als ängstlich. Es wird auch ersehen, daß die am leichtesten erkennbare Kategorie Ärger ist (72,2%) und die am

wenigsten erkennbare Kategorie Angst ist (49,5%). Eine Menge an Verwirrung wird zwischen Traurigkeit und Angst, Traurigkeit und unemotionalem Zustand und Fröhlichkeit und Angst gefunden. Die mittlere Genauigkeit ist 63,5%, welche mit den Resultaten der anderen experimentellen Studien übereinstimmt.

Tabelle 1 Leistungs-Konfusions- bzw. -Verwirrungsmatrix

Tabelle 2 zeigt Statistiken für Bewerter für jede emotionelle Kategorie und für einezusammengefaßte Leistung bzw. Darbietung, welche als die Summe von Darbietungen für jede Kategorie berechnet wurde. Es kann ersehen werden, daß die Varianz bzw. Abweichung für Ärger und Traurigkeit viel weniger als für die anderen emotionellen bzw. emotionalen Kategorien ist.

Tabelle 2 Statistik der Bewerter Merkmalsatz und 10-Neuronen-Architektur auf 53% für die 10-Merkmal- und 10-Neuronen-Architektur. Die Genauigkeit für Fröhlichkeit, Ärger und Traurigkeit ist relativ hoch (68–83%). Die Durchschnittsgenauigkeit (~61%) ist höher als für den s70-Datensatz.

5 zeigt Resultate für einen s90-Datensatz. Wir können sehen, daß die Genauigkeit für Angst höher ist (25–60%), jedoch demselben Muster folgt, wie es für den s80-Datensatz gezeigt wird. Die Genauigkeit für Traurigkeit und Ärger ist sehr hoch: 75–100% für Ärger und 88–93% für Trauer. Die Durchschnittsgenauigkeit (62%) ist ungefähr gleich der Durchschnittsgenauigkeit für den s80-Datensatz.

6 illustriert eine Ausführungsform der vorliegenden Erfindung, welche Emotion unter Verwendung von Statistiken bzw. statistischen Daten detektiert. Zuerst wird eine Datenbank in Vorgang 600 zur Verfügung gestellt. Die Datenbank weist Statistiken auf, welche Statistiken von menschlichen Assoziationen von Stimmparametern mit Emotionen beinhalten, wie beispielsweise jene, die in den Tabellen oben und 3 bis 5 gezeigt sind bzw. werden. Weiters kann die Datenbank eine Serie von Stimmtonhöhen beinhalten, welche mit

Angst assoziiert sind, und eine andere Serie von Stimmtonlagen bzw. -höhen, welche mit Fröhlichkeit assoziiert wird, und einen Fehlerbereich für bestimmte Tonlagen. Als nächstes wird ein Stimmsignal in Vorgang bzw. Funktion 602 empfangen bzw. aufgenommen. In Vorgang 604 wird bzw. werden ein oder mehrere Merkmal(e) aus dem Stimmsignal extrahiert. Siehe den Merkmalsextraktionsabschnitt oben für mehr Details über ein Extrahieren von Merkmalen aus einem Stimmsignal. Dann wird in Vorgang 606 das extrahierte Stimmerkmal mit den Stimmparametern in der Datenbank verglichen. In Vorgang 608 wird eine Emotion aus der Datenbank basierend auf dem Vergleich des extrahierten Stimmerkmals mit den Stimmparametern ausgewählt. Dies kann beispielsweise ein Vergleichen digitalisierter Sprachproben aus der Datenbank mit einer digitalisierten Probe des Merkmals, welches aus dem Stimmsignal extrahiert wurde, um eine Liste von wahrscheinlichen bzw. möglichen Emotionen zu erstellen, und dann ein Verwenden von Algorithmen beinhalten, um Statistiken der Genauigkeit von Menschen beim Erkennen der Emotion zu berücksichtigen, um eine endgültige Bestimmung der wahrscheinlichsten Emotion vorzunehmen. Die ausgewählte Emotion wird endgültig in Vorgang 610 ausgegeben. Siehe den Abschnitt mit dem Titel "Beispielhafte Vorrichtungen zum Detektieren von Emotion in Sprachsignalen", unten, für computerisierte Mechanismen, um eine Emotionserkennung in einer Sprache durchzuführen.

In einer Ausführungsform der vorliegenden Erfindung beinhaltet die Datenbank Wahrscheinlichkeiten von besonderen Merkmalen, welche mit einer Emotion assoziiert werden. Vorzugsweise beinhaltet die Auswahl der Emotion aus der Datenbank ein Analysieren der Wahrscheinlichkeiten und ein Auswählen der wahrscheinlichsten Emotion basierend auf den

Wahrscheinlichkeiten. Optional können die Wahrscheinlichkeiten der Datenbank Darbietungsverwirrungs-Statistiken beinhalten, wie sie beispielsweise in der Bearbeitungs-Verwirrungs-Matrix oben gezeigt sind. Ebenfalls optional können die Statistiken in der Datenbank Selbsterkennungsstatistiken beinhalten, wie sie beispielsweise in den Tabellen oben gezeigt werden.

In einer anderen Ausführungsform der vorliegenden Erfindung beinhaltet das Merkmal, welches extrahiert wird, einen Maximalwert einer Grundfrequenz, eine Standardabweichung der Grundfrequenz, einen Bereich der Grundfrequenz, einen Mittelwert der Grundfrequenz, einen Mittelwert einer Bandbreite einer ersten Formanten, einen Mittelwert einer Bandbreite einer zweiten Formanten, eine Standardabweichung der Energie, eine Sprechgeschwindigkeit bzw. -rate, eine Steigung der Grundfrequenz, einen Maximalwert der ersten Formanten, einen Maximalwert der Energie, einen Bereich der Energie, einen Bereich der zweiten Formanten und/oder einen Bereich der ersten Formanten.

7 ist ein Flußdiagramm, welches ein Verfahren zum Detektieren von Nervosität in einer Stimme in einer Geschäftsumgebung illustriert, um Betrug verhindern zu helfen. Zuerst werden in Vorgang 700 Stimmsignale von einer Person während eines Geschäftsereignisses empfangen. Beispielsweise können die Stimmsignale durch ein Mikrofon in der Nähe der Person erzeugt werden, können von einer Telefonanzapfung eingefangen werden, usw. Die Stimmsignale werden während des

Tabelle drei unten zeigt Statistiken für "Schauspieler", d.h., wie gut Subjekte Emotionen porträtieren. Genauer gesagt, zeigen die Zahlen in der Tabelle, welcher Abschnitt von porträtierten Emotionen einer bestimmten Kategorie als diese Kategorie durch andere Subjekte erkannt wurde. Es ist interessant zu sehen, daß bei einem Vergleich von Tabelle 2 und 3 die Fähigkeit, Emotionen zu porträtieren (Gesamtmittelwert ist 62,9%) ungefähr auf demselben Pegel bleibt wie die Fähigkeit, Emotionen zu erkennen (Gesamtmittelwert ist 63,2%), jedoch die Abweichung bzw. Varianz für ein Porträtieren viel größer ist.

Tabelle 3 Statistik der Schauspieler

Tabelle 4 zeigt Selbstbezugsstatistiken, d.h. wie gut Subjekte fähig waren, ihre eigenen Portraits bzw. Darstellungen zu erkennen. Wir können sehen, daß Leute viel besser ihre eigene Emotionen erkennen (das Mittel ist 80,0%), besonders für Ärger (98,1%), Traurigkeit (80,0%) und Angst (78,8%). Interessanterweise wurde Angst besser erkannt als Glücklichkeit bzw. Fröhlichkeit. Einige Subjekte versagten beim Erkennen ihrer eigenen Darstellungen für Fröhlichkeit und den Normalzustand.

Tabelle 4 Selbstbezugsstatistik

Aus dem Bestand von 700 Aussagen bzw. Äußerungen können fünf ineinander geschachtelte Datensätze, welche Äußerungen beinhalten, die als die gegebene Emotion porträtierend durch wenigstens p Prozent der Subjekte ($p = 70, 80, 90, 95$ und 100%) erkannt wurden, ausgewählt werden. Für die gegenwärtige Besprechung bzw.

Diskussion sollen diese Datensätze als s70, s80, s90 und s100 bezeichnet werden. Tabelle 5 unten zeigt die Anzahl von Elementen in jedem Datensatz. Wir können sehen, daß nur 7,9% der Äußerungen des Bestands durch alle Subjekte erkannt wurden. Und diese Zahl nimmt geradlinig bis zu 52,7% für den Datensatz s70 zu, welcher mit dem 70%-Pegel einer Konkordanz bzw. Übereinstimmung bei einem Dekodieren von Emotionen in Sprache übereinstimmt.

Tabelle 5 p-Pegel Übereinstimmungs-Datensätze

Die Ergebnisse bieten eine nützliche Einsicht über menschliche Darbietung bzw. Leistung und können als eine Grundlinie für einen Vergleich mit einer Computerdarbietung bzw. -leistung dienen.

Merkmalsextraktion

Es wurde gefunden, daß die Tonhöhe der Hauptstimmhinweise zur Emotionserkennung ist. Streng gesprochen, wird die Tonhöhe durch die fundamentale bzw. Grundfrequenz (FO) repräsentiert, d.h. die Haupt-(niedrigste)Frequenz der Vibration bzw. Schwingung der Stimmlippen bzw. Stimmbänder. Die anderen akustischen Variablen, welche zur stimmlichen Emotionssignalisierung beitragen, sind:

- * • Stimmenergie bzw. Vokalenergie
- * • spektrale Frequenz-Merkmale
- * • Formanten (üblicherweise werden nur ein oder zwei erste Formanten (F1, F2) betrachtet).
- * • zeitliche Merkmale (Sprachtempo und Unterbrechung).

Eine andere Annäherung an eine Merkmalsextraktion ist, den Satz von Merkmalen durch Betrachten einiger derivativer Merkmale, wie beispielsweise LPC (lineare, voraussagende, codierende) Parameter eines Signals oder Merkmale der geglätteten Tonhöhenkontur und ihrer Ableitungen zu• Die Erprobung kann auf nur einem Stimmsignal beruhen. Dies bedeutet, daß die betrachten.

Für diese Erfindung kann die folgende Strategie angewendet werden. Erstens, berücksichtige die Grundfrequenz F0 (d.h. die Haupt-(niedrigste)Frequenz der Vibration der Stimmbänder), Energie, Sprachgeschwindigkeit, die ersten drei Formanten (F1, F2 und F3) und ihre Bandbreiten (BW1, BW2 und BW3) und berechne für diese so viele Statistiken bzw. statistische Daten wie möglich. Dann reihe die Statistiken unter Verwendung von Merkmalsauswahltechniken, und wähle einen Satz von "wichtigsten" Merkmalen aus.

Die Sprachgeschwindigkeit kann als das Umgekehrte der Durchschnittslänge des stimmhaften Teils einer Äußerung berechnet werden. Für alle anderen Parameter können die folgenden statistischen Daten berechnet werden: Mittelwert, Standardabweichung, Minimum, Maximum und Bereich. Zusätzlich kann für F0 die Steigung als eine lineare Regression für den stimmhaften Teil der Sprache berechnet werden, d.h. jene Linie, welche zu der Tonhöhenkontur paßt. Die relative stimmhafte bzw. durch Stimme geäußerte Energie kann auch als der Anteil der stimmhaften Energie zu der Gesamtenergie der Äußerung berechnet werden. Insgesamt gibt es etwa 40 Merkmale für jede Äußerung.

Der RELIEF-F- bzw. ENTLASTE-F-Algorithmus kann zur Merkmalsauswahl verwendet werden. Beispielsweise kann ENTLASTE-F für den s70-Datensatz gelaufen werden, wobei die Anzahl der nächsten Nachbarn von 1 bis 12 variiert wird, und die Merkmale entsprechend ihrer Summe von Rängen bzw. Reihungen geordnet werden. Die obersten 14 Merkmale sind die folgenden: F0 Maximum, F0 Standardabweichung, F0 Bereich, F0 Mittelwert, BW1 Mittelwert, BW2 Mittelwert, Energie-Standardabweichung, Sprachgeschwindigkeit, F0 Steigung, F1 Maximum, Energiemaximum, Energiebereich, F2 Bereich und F1 Bereich.

Um zu untersuchen, wie Sätze von Merkmalen die Genauigkeit von Emotionserkennungs- Algorithmen beeinflussen, können drei ineinander verschachtelte Sätze von Merkmalen basierend auf ihrer Summe von Reihungen ausgebildet werden. Der erste Satz beinhaltet die oberen acht

Merkmale (von F0 Maximum zur Sprechgeschwindigkeit), der zweite Satz erstreckt sich von dem ersten zu zwei weiteren Merkmalen (F0 Steigung und F1 Maximum), und der dritte Satz beinhaltet alle 14 Höchstmerkmale. Mehr Details über den RELIEF-F-Algorithmus werden in der Veröffentlichung Proc. European Conf. On Machine Learning (1994) in

dem Artikel von I. Kononenko, mit dem Titel "Abschätzungsattribute: Analyse und Erweiterung von "RELIEF-" bzw. "ENTLASTUNG" dargelegt und auf den Seiten 171–182 gefunden.

2 illustriert eine Ausführungsform der vorliegenden Erfindung, welche Emotion unter Verwendung von Sprach- bzw. -Stimmanalyse detektiert. Im Vorgang 200 wird ein Stimmsignal empfangen bzw. aufgenommen, wie beispielsweise durch ein Mikrophon oder in der Form einer digitalisierten Probe bzw. Abtastung. Eine vorbestimmte Anzahl von Merkmalen des Stimmsignals wird, wie oben dargelegt, extrahiert und in dem Vorgang 202 ausgewählt. Diese Merkmale beinhalten, sind jedoch nicht begrenzt auf, einen Maximalwert einer fundamentalen bzw. Grundfrequenz, eine Standardabweichung der Grundfrequenz, einen Bereich der Grundfrequenz, einen Mittelwert der Grundfrequenz, einen Mittelwert einer Bandbreite einer ersten Formanten, einen Mittelwert einer Bandbreite einer zweiten Formanten, eine Standardabweichung der Energie, eine Sprachgeschwindigkeit, eine Steigung der Grundfrequenz, einen Maximalwert der ersten Formanten, einen Maximalwert der Energie, einen Bereich der Energie, einen Bereich der zweiten Formanten, und einen Bereich der ersten Formanten. Unter Verwendung der in Funktion bzw. Vorgang 202 ausgewählten Merkmale wird eine Emotion, die mit dem Stimmsignal assoziiert ist, in Vorgang 204 basierend auf dem extrahierten Merkmal bestimmt. Schließlich wird in Vorgang 206 die bestimmte Emotion ausgegeben. Siehe die Besprechung unten, besonders unter Bezugnahme auf 8 und 9 für eine detailliertere Besprechung bzw. Diskussion eines Bestimmens einer Emotion basierend auf einem Stimmsignal in Übereinstimmung mit der vorliegenden Erfindung.

Vorzugsweise wird das Merkmal des Stimmsignals aus der Gruppe von Merkmalen ausgewählt, welche aus dem Maximalwert der Grundfrequenz, der Standardabweichung der Grundfrequenz, dem Bereich der Grundfrequenz, dem Mittelwert der Grundfrequenz, dem Mittelwert der Bandbreite Standardabweichung der Energie, und der Sprechgeschwindigkeit bestehen. Idealerweise beinhaltet das extrahierte Merkmal wenigstens eine der Steigung der Grundfrequenz und des Maximalwerts der ersten Formanten.

Optional ist bzw. wird eine Vielzahl von Merkmalen extrahiert, beinhaltend den Maximalwert der Grundfrequenz, die Standardabweichung der Grundfrequenz, den Bereich der Grundfrequenz, den Mittelwert der Grundfrequenz, den Mittelwert der Bandbreite der ersten Formanten, den Mittelwert der Bandbreite der zweiten Formanten, die Standardabweichung der Energie, und die Sprechgeschwindigkeit. Vorzugsweise beinhalten die extrahierten Merkmale die Steigung der Grundfrequenz und den Maximalwert der ersten Formanten.

Als eine andere Option wird eine Vielzahl von Merkmalen extrahiert, beinhaltend den Maximalwert der Grundfrequenz, die Standardabweichung der Grundfrequenz, den Bereich der Grundfrequenz, den Mittelwert der Grundfrequenz, den Mittelwert der Bandbreite der ersten Formanten, den Mittelwert der Bandbreite der zweiten Formanten, die Standardabweichung der Energie, die Sprechgeschwindigkeit, die Steigung der Grundfrequenz, den Maximalwert der ersten Formanten, den Maximalwert der Energie, den Bereich der Energie, den Bereich der zweiten Formanten, und den Bereich der ersten Formanten.

Computerleistung

Um Emotionen in einer Sprache zu erkennen, können zwei beispielhafte Annäherungen vorgenommen werden: neurale Netzwerke und Ensembles von Sortierern bzw.

Klassifiziermaschinen. In der ersten Annäherung kann eine zweilagige rückwärts ausbreitende neurale Netzwerkarchitektur mit einem 8-, 10- oder 14-Element-Eingabevektor, 10 oder 20 Knoten in der versteckten sigmoidalen Schicht und fünf Knoten in der ausgegebenen linearen Schicht verwendet werden. Die Anzahl von Ausgaben stimmt mit der Anzahl von emotionalen Kategorien überein. Um die Algorithmen zu trainieren und zu testen, können Datensätze s70, s80 und s90 verwendet werden. Diese Sätze können zufällig in Training (67% an Äußerungen) und Test (33%) Untersätze aufgespalten bzw. aufgeteilt werden. Verschiedene neurale Netzwerkklassifizierer, welche mit unterschiedlichen Ausgangsgewichts-Matrizen trainiert sind, können erstellt bzw. erzeugt werden. Diese Annäherung, wenn an den s70-Datensatz und den 8-Merkmalsatz oben angewendet, ergaben die Durchschnittsgenauigkeit von etwa 55% mit der folgenden Verteilung für emotionale Kategorien. Normalzustand ist 40–50%, Fröhlichkeit ist 55–65%, Ärger ist 60–80%, Traurigkeit ist 60–70%, und Angst ist 20–40%.

Für die zweite Annäherung bzw. den zweiten Zugang werden Ensembles von Klassifizierern verwendet. Ein Ensemble besteht an einer ungeraden Anzahl von neuronalen Netzwerkklassifizierern, welche auf unterschiedlichen Subsätzen bzw. Untersätzen des Trainingssatzes unter Verwendung der Bootstrapaggregation und kreuz- bzw. querbestätigten Ausschubtechniken trainiert wurden. Das Ensemble fällt Entscheidungen basierend auf dem Mehrheitsabstimmungsprinzip vorgeschlagene bzw. empfohlene Ensemblegrößen reichen von 7 bis 15.

3 zeigt die durchschnittliche Genauigkeit einer Erkennung für einen s70-Datensatz, alle drei Sätze der Merkmale, und beide neurale Netzwerkarchitekturen (10 und 20 Neuronen in der versteckten Schicht). Es kann gesehen werden, daß die Genauigkeit für Glücklichkeit dieselbe bleibt (ungefähr 68%) für die unterschiedlichen Sätze von Merkmalen und Architekturen. Die Genauigkeit für Angst ist ziemlich niedrig (15–25%). Die Genauigkeit für Ärger ist relativ niedrig (40–45%) für den 8-Merkmal-Satz und verbessert sich dramatisch (65%) für den 14-Merkmalssatz. Jedoch ist die Genauigkeit für Traurigkeit höher für den 8-Merkmal-Satz als für die anderen Sätze. Die Durchschnittsgenauigkeit beträgt etwa 55%. Die niedrige Genauigkeit für Angst bestätigt das theoretische Resultat, welches besagt, daß, wenn die individuellen Klassifizierer unkorrelierte Fehler begehen oder Raten 0,5 überschreiten (sie beträgt 0,6–0,8 in unserem Fall), dann die Fehlerrate des gewählten Ensembles zunimmt.

4 zeigt Resultate für einen s80-Datensatz. Es wird gesehen, daß die Genauigkeit für den Normalzustand niedrig ist (20–30%). Die Genauigkeit für Angst ändert sich dramatisch von 11% für Geschäftsereignisses in Vorgang bzw. Funktion 702 analysiert, um einen Nervositätspegel der Person zu bestimmen. Die Stimmsignale können analysiert werden, wie dies oben dargelegt wurde. In Vorgang 704 wird eine Anzeige des Pegels bzw. Niveaus der Nervosität bzw. eines Nervositätspegels ausgegeben, vorzugsweise bevor das Geschäftsereignis abgeschlossen ist, so daß jemand, der versucht, einen Betrug zu verhindern, eine Bewertung vornehmen kann, ob die Person zu konfrontieren ist, bevor diese Person weggeht. Jede Art von Ausgabe ist akzeptierbar, beinhaltend einen Papierausdruck oder eine Anzeige auf einem Computerbildschirm. Es sollte sich verstehen, daß diese Ausführungsform der Erfindung Emotionen verschieden von Nervosität detektieren kann. Derartige Emotionen beinhalten Streß und jede andere Emotion, welche einer Person eigen ist, wenn sie einen Betrug begeht.

Diese Ausführungsform der vorliegenden Erfindung hat eine besondere Anwendung in Geschäftsbereichen, wie beispielsweise Vertragsverhandlung, Versicherungsabwicklungen, Kundenservice, usw. Betrug in diesen Bereichen kostet Gesellschaften jedes Jahr Millionen. Glücklicherweise stellt die vorliegende Erfindung ein Werkzeug zur Verfügung, um bei der Bekämpfung eines derartigen Betrugs zu helfen. Es sollte auch beachtet werden, daß die vorliegende Erfindung Anwendungen im Strafverfolgungsbereich, wie auch in einer Gerichtssaalumgebung usw. hat.

Vorzugsweise wird ein Grad an Gewißheit betreffend den Nervositätspegel der Person ausgegeben, um jemanden bei einer Suche nach Betrug dabei zu helfen, eine Bestimmung darüber vorzunehmen, ob die Person in betrügerischer Absicht gesprochen hat. Dies kann auf Statistiken basieren, wie dies oben in der Ausführungsform der vorliegenden Erfindung unter Bezugnahme auf 6 dargelegt wurde. Optional kann die Anzeige des Nervositätspegels der Person in Echtzeit ausgegeben werden, um es jemanden, der versucht, Betrug zu verhindern, zu gestatten, Resultate sehr schnell zu erhalten, so daß er oder sie fähig ist, die Person bald herauszufordern, nachdem die Person eine verdächtige Äußerung macht.

Als eine andere Option kann die Anzeige des Nervositätspegels einen Alarm beinhalten, welcher ausgelöst wird, wenn der Nervositätspegel einen vorbestimmten Pegel überschreitet. Der Alarm kann eine sichtbare Benachrichtigung auf einer Computeranzeige beinhalten, einen hörbaren Klang bzw. Ton, usw., um einen Aufseher, den Zuhörer und/oder jemanden zu alarmieren, der nach Betrug sucht. Der Alarm könnte auch mit einer Aufzeichnungsvorrichtung verbunden sein, welche beginnen würde, die Konversation aufzuzeichnen, wenn der Alarm ausgelöst wurde, wenn die Konversation nicht bereits aufgezeichnet wird.

Die Alarmoptionen wären besonders in einer Situation nützlich sein, wo sich viele Personen beim Sprechen abwechseln. Ein Beispiel wäre in einer Kundenserviceabteilung oder am Telefon eines Kundenservicebeauftragten. Da jeder Kunde an die Reihe kommt, um mit einem Kundenservicebeauftragten zu sprechen, würden Ausführungsformen der vorliegenden Erfindung den Nervositätspegel in der Sprache des Kunden detektieren. Wenn der Alarm ausgelöst wurde, da der Nervositätspegel eines Kunden den vorgeschriebenen Pegel überquerte, könnte der Kundenservicebeauftragte bzw. -verantwortliche durch eine sichtbare Anzeige auf seinem oder ihrem Computerbildschirm, ein blinkendes Licht, usw. benachrichtigt werden. Der Kundenservicebeauftragte, nun von dem möglichen Betrug wissend, könnte dann versuchen, den Betrug freizulegen bzw. aufzudecken, wenn er existiert. Der Alarm könnte auch verwendet werden, um genauso gut einen Manager zu benachrichtigen. Weiters könnte eine Aufzeichnung der Konversation bzw. Unterhaltung beginnen, nachdem der Alarm aktiviert wurde.

In einer Ausführungsform der vorliegenden Erfindung wird wenigstens ein Merkmal der Stimmsignale extrahiert und verwendet, um den Nervositätspegel der Person zu bestimmen. Merkmale, welche extrahiert werden, können beinhalten einen Maximalwert einer Grundfrequenz, eine Standardabweichung der Grundfrequenz, einen Bereich der Grundfrequenz, einen Mittelwert der Grundfrequenz, einen Mittelwert einer Bandbreite einer ersten Formanten, einen Mittelwert einer Bandbreite einer zweiten Formanten, eine Standardabweichung der Energie, eine Sprechgeschwindigkeit, eine Neigung bzw. Steigung der Grundfrequenz, einen Maximalwert der ersten Formanten, einen Maximalwert der Energie, einen Bereich der Energie, einen Bereich der Formanten, und einen Bereich der ersten Formanten. Somit kann beispielsweise ein Grad eines Schwankens im Ton der Stimme, wie er aus Messungen bzw. Auslesungen der Grundfrequenz bestimmt wurde, verwendet werden, um beim Bestimmen eines Nervositätspegels bzw. -niveaus zu helfen. Je größer der Grad eines Schwankens, umso höher ist der Nervositätspegel. Pausen in der Sprache der Person können ebenfalls berücksichtigt werden.

Der folgende Abschnitt beschreibt Vorrichtungen, welche verwendet werden können, um Emotion beinhaltend Nervosität, in Stimmsignalen zu bestimmen. Beispielhafte Vorrichtungen zum Detektieren von Emotion in Sprach- bzw. Stimmsignalen

Dieser Abschnitt beschreibt verschiedene Vorrichtungen zum Analysieren von Sprache in Übereinstimmung mit der vorliegenden Erfindung.

Eine Ausführungsform der vorliegenden Erfindung beinhaltet eine Vorrichtung zum Analysieren der Sprache einer Person, um ihren emotionalen Zustand zu bestimmen. Der Analysator arbeitet auf der Echtzeitfrequenz oder Tonhöhenkomponenten innerhalb des ersten Formantenbands von menschlicher Sprache. Beim Analysieren der Sprache analysiert die Vorrichtung Erscheinungsmuster bestimmter Werte im Hinblick auf Muster einer differentiellen ersten Formantentonlage, Rate einer Tonlagenänderung, Dauer und Zeitverteilung. Diese Faktoren beziehen sich auf eine komplexe, jedoch sehr fundamentale Weise sowohl auf vorübergehende wie auch emotionellen Langzeit-Zuständen.

Die menschliche Sprache wird durch zwei grundlegende Tonerzeugungsmechanismen angeregt. Die Stimmbänder; dünne gestreckte bzw. gedehnte Membranen unter Muskelregelung bzw. -steuerung oszillieren bzw. schwingen, wenn ausgestoßene Luft von den Lungen durch sie durchtritt. Sie erzeugen einen charakteristischen "Brumm"-Klang bei einer fundamentalen bzw. Grundfrequenz zwischen 80 Hz und 240 Hz. Diese Frequenz wird über einen moderaten Bereich sowohl durch bewußte als auch unbewußte Muskelkontraktion und -entspannung variiert. Die Wellenform des grundlegenden "Brummens" beinhaltet viele Harmonische, von welchen einige eine Erregungsresonanz verschieden festgelegt ist und veränderliche Hohlräume mit dem stimmlichen Gebiet assoziiert sind. Der zweite Grundton, welcher während einer Sprache erzeugt wird, ist ein pseudozufälliges Rauschen, welches eine ziemlich breite und einheitliche bzw. gleichförmige Frequenzverteilung aufweist. Er wird durch Turbulenz verursacht, sobald ausgeatmete bzw. ausgestoßene Luft sich durch den Vokaltrakt bewegt und wird ein "Zisch"-Klang bzw. Zischlaut genannt. Er wird hauptsächlich durch Zungenbewegungen moduliert und erregt auch die festgelegten und veränderlichen Hohlräume. Es ist diese komplexe Mischung von "Brumm"- und "Zisch"-Lauten, welche durch die Resonanzhohlräume geformt und artikuliert werden, welche Sprache erzeugen.

In einer Energieverteilungsanalyse von Sprachklängen bzw. Sprachtönen wird gefunden werden, daß die Energie in getrennte bzw. ausgeprägte Frequenzbänder, Formanten genannt, fällt. Es gibt drei signifikante Formanten. Das hier beschriebene System verwendet das erste Formantenband, welches sich von der grundlegenden "Brumm"-Frequenz bis ungefähr 1000 Hz erstreckt. Dieses Band weist nicht nur den höchsten Energiegehalt auf, sondern reflektiert einen hohen Grad an Frequenzmodulation als eine Funktion von verschiedenen Vokaltrakt- und Gesichtsmuskelspannungs-Variationen.

In Wirklichkeit wird durch ein Analysieren bestimmter Verteilungsmuster der ersten Formantenfrequenz eine qualitative Messung von sprachbezogenen Muskelspannungsvariationen und Wechselwirkungen durchgeführt. Da diese Muskel überwiegend durch sekundäre unbewußte Vorgänge vorgespannt und artikuliert werden, welche wiederum durch einen emotionalen Zustand beeinflusst sind, kann eine relative Messung einer emotionalen Aktivität unabhängig von dem Bewußtsein einer Person oder einem Mangel an Bewußtsein von diesem Zustand bestimmt werden. Die Forschung bestätigt auch eine allgemeine Vermutung, daß, da die Mechanismen der Sprache äußerst komplex und weitestgehend autonom sind, sehr wenig Leute fähig sind,

bewußt einen fiktiven emotionalen Zustand zu "projizieren". Tatsächlich erzeugt ein Versuch so vorzugehen, üblicherweise seinen eigenen einzigartigen psychologischen Streß-"Fingerabdruck" in dem Stimmuster.

Wegen der Charakteristika bzw. Merkmale der ersten Formanten-Sprachklänge bzw. -Sprachtöne analysiert die vorliegende Erfindung ein FM-demoduliertes erstes Formanten-Sprachsignal und erzeugt eine Ausgabe, welche Nullen davon anzeigt.

Die Frequenz oder Anzahl von Nullen oder "flachen" Punkten in dem FM-demodulierten Signal, die

Länge der Nullen und das Verhältnis der Gesamtzeit, zu welcher Nullen während einer Wortperiode existieren, zur Gesamtzeit der Wortperiode sind alle für einen emotionalen Zustand des Einzelnen anzeigend bzw. indikativ. Durch ein Betrachten der Ausgabe der Vorrichtung kann der Ver- bzw. Anwender das Auftreten der Nullen sehen oder fühlen und somit durch ein Beobachten der Ausgabe der Anzahl oder Frequenz von Nullen bestimmen, der Länge der Nullen und des Verhältnisses der Gesamtzeit, während welcher Nullen während einer Wortperiode bis zur Länge der Wortperiode existieren, den emotionalen Zustand des Einzelnen bzw. Individiums.

In Ausführungsformen der vorliegenden Erfindung ist das erste Formanten-Frequenzband eines Sprachsignals FM-demoduliert und das FM-demodulierte Signal wird an einen Wortdetektorschaltkreis angewendet bzw. angelegt, welcher das Vorhandensein eines FM-demodulierten Signals detektiert. Das FM-demodulierte Signal wird auch an Null-Detektormittel angelegt, welche die Nullen in dem FM-demodulierten Signal detektieren, und eine Ausgabe erzeugen, welche dafür anzeigend bzw. hinweisend ist. Ein Ausgabeschaltkreis wird mit dem Wortdetektor und dem Nulldetektor gekoppelt. Der Ausgabeschaltkreis wird durch den Wortdetektor freigegeben, wenn der Wortdetektor das Vorhandensein eines FM-demodulierten Signals detektiert, und der Ausgabeschaltkreis erzeugt eine Ausgabe, welche für das Vorhandensein oder Nicht-Vorhandensein einer Null in dem FM-demodulierten Signal anzeigend ist. Die Ausgabe des Ausgabeschaltkreises wird auf eine Weise angezeigt, in welcher sie durch einen Anwender wahrgenommen wird, so daß der Ver- bzw. Anwender mit einer Anzeige des Bestehens von Nullen in dem FM-demodulierten Signal versorgt ist. Der Anwender der Vorrichtung überwacht somit die Nullen und kann dadurch den emotionalen Zustand des Individuums bestimmen, dessen Sprache analysiert wird.

In einer anderen Ausführungsform der vorliegenden Erfindung wird das Stimmvibrato analysiert. Das sogenannte Stimmvibrato wurde als eine halbfreiwillige Antwort erstellt, welche beim Studieren einer Irreführung zusammen mit bestimmten anderen Reaktionen von Wert sein könnte; wie beispielsweise Atmungsvolumen; Einatmungs-Ausatmungs-Verhältnisse; Stoffwechselrate; Regelmäßigkeit und Rate bzw. Geschwindigkeit einer Einatmung; Assoziation von Worten und Ideen; Gesichtsausdrücke; Bewegungsreaktionen; und Reaktionen auf bestimmte Narkotika; jedoch wurde keine verwendbare Technik zuvor entwickelt, welche eine gültige und zuverlässige Analyse von Stimmänderungen in der klinischen Bestimmung eines emotionalen Zustands, Meinungen oder Täuschungsversuche eines Subjekts bzw. einer Person gestattet.

Frühe Experimente, welche Versuche beinhalteten, Stimmqualitätsänderungen mit emotionalen Reizen zu korrelieren, haben festgestellt bzw. ergeben, daß die menschliche Sprache durch starke Emotion beeinflusst wird. Detektierbare Änderungen der Stimme treten viel rascher auf, nachfolgend auf eine Streßstimulierung, als es die klassischen Anzeigen von physiologischen Manifestationen tun, welche aus dem Funktionieren des autonomen Nervensystems resultieren.

Zwei Typen einer Stimme ändern sich als ein Resultat von Streß. Die erste von diesen wird als die Grobänderung bezeichnet, welche üblicherweise nur als ein Ergebnis einer wesentlichen Streßsituation auftritt. Diese Änderung manifestiert sich selbst in hörbaren merklichen Änderungen der Sprechgeschwindigkeit, Lautstärke, Stimmzittern, Änderung im Abstand zwischen Silben, und einer Änderung in der Grundtonhöhe oder Frequenz der Stimme. Diese grobe Änderung ist Gegenstand der bewußten Steuerung, wenigstens bei einigen Subjekten bzw. Personen, wenn der Streßpegel unter jenem eines Totalverlusts einer Steuerung bzw. Regelung ist.

Der zweite Typ an Stimmänderung ist jener von Stimmqualität. Dieser Typ der Änderung ist nicht für das menschliche Ohr unterscheidbar bzw. wahrnehmbar, sondern ist eine offenbar unbewußte Manifestation der leichten Spannung der Stimmbänder unter sogar geringem Streß, resultierend in einer Dämpfung ausgewählter Frequenzvariationen. Bei graphischer Darstellung wird der

Unterschied leicht zwischen ungestreßter oder normaler Stimmgebung und Stimmgebung unter mildem bzw. geringem Streß, Täuschungsversuchen oder

feindlichen bzw. gegnerischen Einstellungen wahrnehmbar. Diese Muster haben sich über einen weiten Bereich menschlicher Stimmen beider Geschlechter, unterschiedlicher Alter und unter verschiedenen situationalen Bedingungen als wahr bzw. zutreffend erwiesen. Dieser zweite Änderungstyp ist nicht Gegenstand einer bewußten Steuerung.

Es gibt zwei Typen bzw. Arten von Klang bzw. Ton, welche durch die menschliche Stimmanatomie erzeugt werden. Der erste Typ an Ton ist ein Produkt aus der Vibration der Stimmbänder, welche wiederum ein Produkt eines teilweisen Schließens der Stimmritze ist und eines Zwingens von Luft durch die Stimmritze durch eine Kontraktion des Lungenhohlraums und der Lungen ist. Die Frequenzen dieser Vibrationen können im allgemeinen zwischen 100 und 300 Hertz, abhängig von Geschlecht und Alter des Sprechers und den Intonationen variieren, die der Sprecher anwendet. Dieser Ton hat eine rasche Abfall- bzw. Abklingzeit.

Die zweite Art an Klang bzw. Ton beinhaltet die Formanten-Frequenzen. Diese bilden einen Klang, welcher aus der Resonanz der Hohlräume im Kopf, beinhaltend den Hals, den Mund, die Nase und die Stirnhöhlen resultiert. Der Klang wird durch eine Anregung der Resonanzhöhlräume durch eine Tonquelle von niedrigeren Frequenzen, im Fall des durch die Stimmbänder erzeugten vokalisiertes Klangs, oder durch eine teilweise Beschränkung des Durchtritts von Luft von den Lungen, wie im Fall von stimmlosen Reiblauten erzeugt. Was auch immer die Erregungsquelle ist, die Frequenz der Formanten wird durch die Resonanzfrequenzen des involvierten Hohlraums bestimmt. Die Formantenfrequenzen erscheinen im allgemeinen bei etwa 800 Hertz und erscheinen in bestimmten Frequenzbändern, welche mit der Resonanzfrequenz der individuellen Hohlräume übereinstimmen. Die erste, oder niedrigste, Formante, ist jene, die durch den Mund und Rachen- bzw. Halshohlräume gebildet wird und ist merkbar für ihre Frequenzverschiebung, sobald der Mund seine Abmessungen und Volumen bei der Bildung von verschiedenen Klängen, besonders von Vokalen ändert. Die höchsten Formanten-Frequenzen sind konstanter aufgrund des konstanteren Volumens der Hohlräume. Die Formanten-Wellenformen sind läutende Signale, im Gegensatz zu den rasch abklingenden Signalen der Stimmbänder. Wenn stimmhafte Töne bzw. Klänge geäußert werden, werden die Stimmwellenformen auf die Formanten-Wellenformen als Amplitudenmodulationen eingepreßt bzw. überlagert.

Es wurde entdeckt, daß eine dritte Signalkategorie in der menschlichen Stimme existiert und daß diese dritte Signalkategorie mit der zweiten Art einer Stimmänderung verwandt ist, wie dies oben besprochen wurde. Dies ist eine Infraschall- oder Unterschall-Frequenzmodulation, welche in einem gewissen Grad sowohl in den Stimmbandklängen als auch in den Formantenklängen vorhanden ist. Dieses Signal liegt typischerweise zwischen 8 und 12 Hertz. Dementsprechend ist es nicht für das menschliche Ohr hörbar. Wegen der Tatsache, daß diese Charakteristik eine Frequenzmodulation bildet, wie sie sich von einer Amplitudenmodulation unterscheidet, ist es nicht direkt auf Zeitbasis/Amplituden-Kartenaufzeichnungen wahrnehmbar bzw. unterscheidbar. Wegen der Tatsache, daß dieses Infraschallsignal eine der signifikanteren Stimmanzeigen von psychologischem Streß ist, wird es in größerem Detail behandelt werden.

Es bestehen verschiedene Analogien, welche verwendet werden, um schematische Darstellungen des gesamten Stimmprozesses zur Verfügung zu stellen. Sowohl mechanische wie auch elektronische Analogien werden erfolgreich beispielsweise beim Entwurf von Computerstimmen verwendet bzw. angewandt. Diese Analogien betrachten jedoch die stimmhafte Klangquelle (Stimmbänder) und die Wände der Hohlräume als harte und konstante Merkmale. Jedoch stellen sowohl die Stimmbänder als auch die Wände der grundlegenden Formanten-erzeugenden Hohlräume in der Realität ein flexibles Gewebe dar, welches augenblicklich auf die komplexe

Anordnung von Muskeln antwortet, welche eine Steuerung des Gewebes zur Verfügung stellen. Diese Muskel, welche die Stimmbänder durch die mechanische Verbindung von Knochen und Knorpel steuern, gestatten sowohl die gezielte als auch automatische Erzeugung von Stimmklang und Veränderung der Stimmtonehöhe durch ein Individuum. In ähnlicher Weise erlauben diese Muskel, welche die Zunge, Lippen und Hals bzw. Rachen steuern, sowohl die gezielte als auch die automatische Steuerung der ersten Formanten-Frequenzen. Andere Formanten können ähnlich zu einem begrenzteren Grad bzw. Ausmaß beeinflusst werden.

Es ist wert zu beachten, daß während normaler Sprache diese Muskel auf einem kleinen Prozentsatz ihrer Gesamtarbeitsfähigkeit arbeiten. Aus diesem Grund verbleiben, trotz ihrer Verwendung zum Ändern der Position der Stimmbänder und der Positionen der Lippen, Zunge und inneren Halswände, die Muskeln in einem relativ entspannten Zustand. Es wurde bestimmt, daß während dieses relativ entspannten Zustands eine Muskelwellenbewegung typischerweise bei der zuvor erwähnten Frequenz von 8 bis 12 Hertz auftritt. Diese Wellenbewegung verursacht eine leichte Variation in der Spannung der Stimmbänder und verursacht Verschiebungen bzw. Verlagerungen der grundlegenden Tonhöhenfrequenz der Stimme. Auch variiert die

Wellenform leicht das Volumen des Resonanzhohlraums (besonders des mit der ersten Formanten assoziierten) und der Elastizität der Hohlraumwände, um Verschiebungen der Formantenfrequenzen zu verursachen. Diese Verschiebungen um eine Zentralfrequenz bilden eine Frequenzmodulation der Zentral- oder Trägerfrequenz.

Es ist wichtig zu beachten, daß keine der Verschiebungen in der grundlegenden Tonhöhenfrequenz der Stimme oder in den Formantenfrequenzen direkt durch einen Zuhörer detektierbar ist, teilweise, da die Verschiebungen sehr klein sind, und teilweise, da sie primär in dem zuvor erwähnten unhörbaren Frequenzbereich bestehen.

Um diese Frequenzmodulation zu beobachten, kann irgendeine von verschiedenen bestehenden Techniken zur Demodulation einer Frequenzmodulation eingesetzt bzw. verwendet werden, natürlich unter Beachtung, daß die Modulationsfrequenz die nominellen 8–12 Hertz sind und der Träger eines der Bänder innerhalb des Stimmspektrums ist.

Um vollständiger die obige Diskussion zu verstehen, muß das Konzept eines "Schwerpunkts" dieser Wellenform verstanden werden. Es ist möglich, ungefähr den Mittelpunkt zwischen den zwei Extremen bzw. Extremwerten von jeder einzelnen Auslenkung des Aufzeichnungsstifts zu bestimmen. Wenn die Mittelpunkte zwischen den Extremen aller Auslenkungen markiert sind und wenn diese Mittelpunkte dann ungefähr durch eine kontinuierliche Kurve verbunden sind, wird gesehen werden, daß eine Linie, die sich einem Mittelwert oder "Schwerpunkt" der gesamten Wellenform annähert, resultieren wird. Ein Verbinden aller derartigen Markierungen, mit etwas Glättung, resultiert in einer sanften gekrümmten Linie. Die Linie repräsentiert die Infraschall-Frequenzmodulation, die aus den zuvor beschriebenen Wellenformen resultiert.

Wie oben erwähnt, wurde bestimmt, daß die Anordnung von mit den Stimmbändern assoziierten Muskeln und Hohlraumwänden Gegenstand einer sanften Muskelspannung ist, wenn leichter bis mittlerer psychologischer Streß in der individuellen Überprüfung bzw. Untersuchung erzeugt wird. Diese Spannung, die für das Subjekt bzw. die Person nicht wahrnehmbar ist, und ähnlich durch normale, nicht unterstützte Beobachtungstechniken für den Überprüfer nicht wahrnehmbar ist, ist ausreichend, um die Muskelwellenformen zu senken oder virtuell die Muskelwellenformen zu verringern oder nahezu zu entfernen, welche in dem nicht gestreßten Subjekt vorhanden sind, wodurch die Basis für die Trägerfrequenzvariationen entfernt wird, welche die Infraschall-Frequenzmodulierungen erzeugen.

Während die Verwendung der Infraschallwellenform einzigartig für die Technik ist, welche Stimme als das physiologische Medium zur physiologischen Streßbewertung einsetzt, stellt die Stimme zusätzliche instrumentierte Anzeigen von durch ein Hören nicht wahrnehmbaren physiologischen Änderungen als ein Ergebnis von psychologischem Streß zur Verfügung, welche physiologischen Änderungen ähnlich durch Techniken und Vorrichtungen in gegenwärtiger Verwendung detektierbar sind. Von den vier am häufigsten verwendeten physiologischen Änderungen, welche zuvor erwähnt wurden (Hirnwellenmuster, Herzaktivität, Hautleitfähigkeit und Atmungsaktivität) beeinflussen zwei von diesen, die Atmungsaktivität und Herzaktivität, direkt und indirekt die Amplitude und das Detail einer Wellenform einer mündlichen Äußerung und stellen die Basis für eine gröbere Bewertung von psychologischem Streß, besonders, wenn die Prüfung bzw. das Testen sequentielle vokale Antworten beinhaltet bzw. involviert, zur Verfügung.

Eine andere Vorrichtung ist in 8 gezeigt. Wie gezeigt, konvertiert bzw. wandelt ein Wandler 800 die Schallwellen der mündlichen Äußerungen des Subjekts in elektrische Signale um, wovon diese mit der Eingabe eines Audioverstärkers 802 verbunden sind, welcher einfach zum Zweck einer Leistungssteigerung der elektrischen Signale auf einen stabileren nützlicheren Pegel dient. Die Ausgabe bzw. der Ausgang des Verstärkers 802 ist mit einem Filter 804 verbunden, welches primär zum Zweck eines Entfernens einiger unerwünschten Niederfrequenzkomponenten und Rauschkomponenten dient.

Nach einem Filtern wird das Signal mit einem FM-Diskriminator 806 verbunden, wobei die Frequenzabweichungen von der zentralen bzw. Mittenfrequenz in Signale umgewandelt werden, welche in der Amplitude variieren. Die in der Amplitude variierenden Signale werden dann in einem Detektorschaltkreis 808 zum Zweck eines Gleichrichtens des Signals detektiert und erzeugen ein Signal, welches eine Serie von Halbwellenpulsen bildet. Nach einer Detektion wird das Signal mit einem Integratorschaltkreis 810 verbunden, worin das Signal im gewünschten Ausmaß integriert wird. Im Schaltkreis bzw. in der Schaltung 810 wird das Signal entweder in einem sehr kleinen Ausmaß integriert, wobei eine Wellenform erzeugt wird, oder wird zu einem größeren Ausmaß integriert, wobei ein Signal erzeugt wird. Nach einer Integration wird das Signal im Verstärker 812 verstärkt und mit einem Prozessor 814 verbunden, welcher die mit dem Stimmsignal assoziierte Emotion bestimmt. Eine Ausgabevorrichtung 816, wie beispielsweise ein Computerbildschirm oder Drucker wird verwendet, um die detektierte Emotion auszugeben. Optional können statistische Daten ebenfalls ausgegeben werden.

Eine etwas einfachere Ausführungsform einer Vorrichtung zum Erzeugen sichtbarer Aufzeichnungen in Übereinstimmung mit der Erfindung ist in 9 gezeigt, worin die akustischen Signale durch ein Mikrophon 900 in elektrische Signale umgewandelt werden, welche magnetisch in einer Bandaufzeichnungsvorrichtung 902 aufgezeichnet werden. Die Signale können dann durch die verbleibende Ausrüstung bei verschiedenen Geschwindigkeiten und zu jeder Zeit be- bzw. verarbeitet werden, wobei die Wiedergabe mit einer herkömmlichen Halbleiterdiode 904 verbunden ist, welche die Signale gleichrichtet. Die gleichgerichteten Signale werden mit der Eingabe eines herkömmlichen bzw. konventionellen Verstärkers 906 verbunden und auch mit dem bewegbaren Kontakt eines Wahlschalters, welcher allgemein bei 908 angezeigt bzw. bezeichnet ist. Der bewegbare Kontakt des Schalters 908 kann zu irgendeinem einer Vielzahl von festgelegten Kontakten bewegt werden, von welchen ein jeder mit einem Kondensator verbunden ist. In 9 ist eine Auswahl von vier Kondensatoren 910, 912, 914 und 916 gezeigt, wovon ein jeder einen Anschluß mit einem festgelegten Kontakt des Schalters verbunden und den anderen Anschluß mit Masse verbunden aufweist. Die Ausgabe bzw. der Ausgang des Verstärkers 906 ist mit einem Prozessor 918 verbunden.

Ein Bandaufzeichnungsgerät, welches in dieser besonderen Anordnung von Ausrüstung verwendet werden kann, war eine Uher Modell-4000 Bändeinheit mit vier Geschwindigkeiten, welche ihren

eigenen internen Verstärker aufweist. Die Werte der Kondensatoren 910 bis 916 waren jeweils 0,5, 3, 10 und 50 Mikrofarad und die Eingabeimpedanz des Verstärkers 906 war ungefähr 10.000 Ohm. Wie erkannt werden wird, könnten verschiedene andere Komponenten in dieser Vorrichtung verwendet werden bzw. verwendet worden sein.

Im Betrieb des Schaltkreises von 9 wird die durch Diode 904 entstehende gleichgerichtete Wellenform im gewünschten Ausmaß integriert, wobei die Zeitkonstante so ausgewählt ist, daß der Effekt der frequenzmodulierten Infrashallwelle als ein langsam variierender Gleichstrom- bzw. DC-Pegel erscheint, welcher ungefähr der Linie folgt, die den "Schwerpunkt" der Wellenform repräsentiert. Die in diesem besonderen Diagramm gezeigten Ausschläge sind relativ rasch, wobei dies anzeigt, daß der Schalter mit einem der Kondensatoren mit niedrigerem Wert verbunden war. In dieser Ausführungsform wird ein gemischtes bzw. zusammengesetztes Filtern durch den Kondensator 910, 912, 914 oder 916, und in dem Fall einer Reduktion der Wiedergabegeschwindigkeit durch das Bandaufzeichnungsgerät erreicht. der ersten Formanten, dem Mittelwert der Bandbreite der zweiten Formanten, der den 8- zweiten

Telefonischer Betrieb mit Bedienerückmeldung

10 illustriert eine Ausführungsform der vorliegenden Erfindung, welche Emotionen in Stimmsignalen überwacht und eine Bedienerückkopplung bzw. Telefonistenrückkopplung bzw. -rückmeldung basierend auf den detektierten Emotionen zur Verfügung stellt. Zuerst wird ein Stimmsignal, welches für eine Komponente einer Konversation zwischen wenigstens zwei Subjekten repräsentativ ist, in Funktion bzw. Vorgang 1000 empfangen. In Vorgang 1002 wird eine Emotion, welche mit dem Stimmsignal assoziiert ist, bestimmt. Schließlich wird in Vorgang 1004 eine Rückkopplung bzw. Rückmeldung einem Dritten basierend auf der bestimmten Emotion zur Verfügung gestellt.

Die Konversation kann über ein Telekommunikationsnetzwerk ausgeführt werden, wie auch ein Weitverkehrsnetzwerk, wie beispielsweise das Internet, wenn es mit Internettelephonie verwendet wird. Als eine Option werden die Emotionen gerastet bzw. gesiebt und eine Rückmeldung wird nur zur Verfügung gestellt, wenn von der Emotion bestimmt wird, eine negative Emotion zu sein, ausgewählt aus der Gruppe von negativen Emotionen, welche aus Ärger, Traurigkeit und Angst bestehen. Das Gleiche könnte mit Gruppen von positiven oder neutralen Emotionen vorgenommen werden. Die Emotion kann durch ein Extrahieren eines Merkmals aus dem Stimmsignal bestimmt werden, wie dies zuvor im Detail beschrieben wurde.

Die vorliegende Erfindung ist besonders für einen Betrieb in Verbindung mit einem Notfallantwortsystem, wie beispielsweise dem 911-System geeignet. In einem derartigen System könnten eingehende Anrufe durch Ausführungsformen der vorliegenden Erfindung überwacht werden. Eine Emotion des Anrufers würde während der Konversation des Anrufers mit dem Techniker bestimmt werden, der den Anruf beantwortet. Die Emotion könnte dann über Funkwellen beispielsweise zu dem Notfallteam gesandt werden, d.h. Polizei, Feuerwehr und/oder Ambulanzpersonal, so daß diese über den emotionellen Zustand des Anrufers unterrichtet sind.

In einem anderen Szenario ist einer der Subjekte ein Kunde, ein anderes der Subjekte ist ein Angestellter, wie beispielsweise jemand, der durch ein Callcenter oder eine Kundenserviceabteilung angestellt ist, und der Dritte oder die dritte Person ist ein Manager. Die vorliegende Erfindung würde die Konversation zwischen dem Kunden und dem Angestellten überwachen, um zu bestimmen, ob der Kunde und/oder der Angestellte beispielsweise aufgeregt wird. Wenn negative Emotionen detektiert werden, wird eine Rückmeldung zum Manager gesandt, der die Situation beurteilen und falls nötig, intervenieren bzw. einschreiten kann.

Verbessern einer Emotionserkennung

11 illustriert ein System, das eine Anwender- mit einer Computeremotionsdetektion von Stimmsignalen vergleicht, um eine Stimmerkennung von einer der Ausführungsformen der Erfindung, einem Anwender oder beiden zu verbessern. Zuerst werden, in Vorgang 1100, ein Stimmsignal und eine Emotion, welche mit dem Stimmsignal assoziiert ist, zur Verfügung gestellt. Die mit dem Stimmsignal assoziierte Emotion wird automatisch in Vorgang 1102 auf eine Weise bestimmt, wie sie oben dargelegt wurde. Die automatisch bestimmte Emotion wird in Vorgang 1104 beispielsweise auf einem computerlesbaren Medium gespeichert. In Vorgang 1106 wird eine anwenderbestimmte, mit dem Stimmsignal assoziierte Emotion, welche durch einen Anwender bestimmt ist, empfangen. Die automatisch bestimmte Emotion wird mit der anwenderbestimmten Emotion in Vorgang 1108 verglichen.

Das Stimmsignal kann ausgesendet bzw. ausgegeben oder empfangen werden durch ein System, welches die vorliegende Erfindung verkörpert. Optional wird die mit dem Stimmsignal assoziierte Emotion identifiziert, wenn bzw. nachdem die Emotion bereitgestellt ist. In einem derartigen Fall sollte bestimmt werden, ob die automatisch bestimmte Emotion oder die anwenderbestimmte Emotion mit der identifizierten Emotion übereinstimmt. Dem Ver- bzw. Anwender kann ein Preis zuerkannt werden, wenn die anwenderbestimmte Emotion mit der identifizierten Emotion übereinstimmt bzw. zusammenpaßt. Weiters kann die Emotion automatisch durch ein Extrahieren Weise, bestimmt werden.

Um einen Anwender beim Erkennen einer Emotion zu unterstützen, kann ein Emotionserkennungsspiel gespielt werden. Das Spiel könnte es einem Anwender gestatten, gegen einen Computer oder eine andere Person anzutreten, um zu sehen, wer am besten Emotion in aufgezeichneter Sprache erkennen kann. Eine praktische Anwendung des Spiels ist, autistischen Leuten bei einer Entwicklung einer besseren emotionalen Erfahrung beim Erkennen von Emotion in der Sprache zu helfen.

In einer Anordnung kann eine Vorrichtung verwendet werden, um Daten über Stimmsignale zu erzeugen, welche verwendet werden können, um eine Emotionserkennung zu verbessern. Die Vorrichtung akzeptiert stimmlichen Klang durch einen Wandler, wie beispielsweise ein Mikrophon oder ein Klang- bzw. Tonaufzeichnungsgerät. Die physikalische Schallwelle bzw. Klangwelle, welche in elektrische Signale umgewandelt wurde, wird parallel an einer typischen, kommerziell verfügbaren Bank von elektronischen Filtern angewendet, welche den hörbaren bzw. Audio-Frequenzbereich abdecken. Ein Einstellen der zentralen bzw. Mittelfrequenz des niedrigsten Filters auf irgendeinen Wert, welcher die elektrische Energierepräsentation der Stimmsignalamplitude durchläßt, welche das niedrigste Stimmfrequenzsignal beinhaltet, errichtet die Mittenwerte aller nachfolgenden Filter bis zum letzten, welcher im allgemeinen die Energie zwischen 8 kHz bis 16 kHz oder zwischen 10 kHz und 20 kHz durchläßt, und auch die exakte Anzahl derartiger Filter bestimmt. Der spezifische Wert der Mittenfrequenz des ersten Filters ist nicht signifikant, solange die tiefsten Töne der menschlichen Stimme aufgefangen werden, ungefähr 70 Hz. Im wesentlichen ist jede kommerziell erhältliche Bank anwendbar, wenn sie an irgendeinen kommerziell erhältlichen Digitalisierer und dann Mikrocomputer angeschlossen werden kann. Der Beschreibungsabschnitt beschreibt einen spezifischen Satz von Mittenfrequenzen und Mikroprozessor in der bevorzugten Ausführungsform. Die Filterqualität ist auch nicht besonders signifikant, da ein in der Beschreibung geoffenbarter Verbesserungs- bzw. Verfeinerungsalgorithmus jeden Satz von Filtern durchschnittlicher Qualität in akzeptable Frequenz- und Amplitudenwerte bringt. Das Verhältnis 1/3 definiert natürlich die Bandbreite von allen Filtern, sobald die Mittenfrequenzen berechnet sind.

Diesem Sequentierungs- bzw. Aufteilungsvorgang bzw. -prozeß folgend werden die Filterausgabespannungen durch einen kommerziell erhältlichen Satz von Digitalisierern oder

vorzugsweise Multiplexer und Digitalisierern oder einen Digitalisierer digitalisiert, welcher in die gleiche identifizierte kommerziell erhältliche Filterbank eingebaut ist, um eine Kopplungslogik und Hardware zu entfernen. Erneut ist die Qualität eines Digitalisierers bzw. einer Digitalisiervorrichtung im Hinblick auf die Geschwindigkeit einer Umwandlung oder Diskriminierung nicht signifikant bzw. bedeutsam, da durchschnittliche gegenwärtig erhältliche kommerzielle

Einheiten die hier benötigten Erfordernisse aufgrund eines Korrekturalgorithmus (siehe Spezifikationen) und der niedrigen notwendigen Abtastrate überschreiten.

Jeder komplexe Klang bzw. Ton bzw. Schall, welcher sich konstant ändernde Information trägt, kann mit einer Reduktion von Informationsbits durch ein Auffangen der Frequenz und Amplitude von Spitzen des Signals angenähert werden. Dies ist natürlich altes Wissen, wie es ein Durchführen eines derartigen Vorgangs an Sprachsignalen auch ist. Jedoch waren in der Sprachforschung verschiedene spezifische Bereiche, wo derartige Spitzen häufig auftreten, als "Formanten"-Bereiche bezeichnet. Jedoch decken sich diese Bereichsannäherungen nicht immer mit den Spitzen jedes Sprechers unter allen Umständen. Sprachforscher und der erfinderische Stand der Technik neigen zu einer großen Anstrengung, um "legitimierte" Spitzen zu messen und als solche zu benennen, welche innerhalb der typischen Formanten-Frequenzbereiche fallen, als ob ihre Definition nicht Abschätzungen involvieren würde, sondern eher Absolutheit bzw. Absolutwerte. Dies hat zahlreiche Forschung und Formanten-Meßvorrichtungen veranlaßt, künstlich entsprechende Spitzen auszuschließen, welche nötig sind, um adäquat eine komplexe, hochveränderliche Klangwelle in Echtzeit zu repräsentieren. Da die vorliegende Offenbarung entworfen ist, um für Tierstimmenklänge wie auch alle menschlichen Sprachen geeignet zu sein, sind künstliche Beschränkungen, wie beispielsweise Formanten, nicht von Interesse, und die Klang- bzw. Schallwelle wird als eine komplexe, variierende Schallwelle behandelt, welche jeden derartigen Klang bzw. Schall analysieren kann.

Um eine Spitzenidentifizierung unabhängig von einer Abweichung in Filterbandbreite, Qualität und Digitalisiererdiskriminierung zu normalisieren und zu vereinfachen, sind die tatsächlichen Werte, welche für Amplitude und Frequenz gespeichert sind, "repräsentative Werte". Dies deshalb, damit die Breite von oberen Frequenzfilter numerisch ähnlich zur Bandbreite der Filter unterer Frequenz ist. Jedem Filter werden einfach fortlaufende Werte von 1 bis 25 gegeben, und ein Klang von sanft zu laut wird von 1 bis 40 zur Vereinfachung auf einer CRT-Bildschirmanzeige skaliert. Eine Korrektur an den Frequenzrepräsentationswerten wird erreicht bzw. durchgeführt, indem die Anzahl der Filter auf einen höheren Dezimalwert zu dem nächsten ganzzahligen Wert eingestellt wird, wenn die Filterausgabe zur Rechten des Spitzenfilters eine größere Amplitude als die Filterausgabe zur Linken des Spitzenfilters aufweist. Die Details dieses Algorithmus sind in den Beschreibungen dieser Offenbarung beschrieben. Dieser Korrekturvorgang muß sich vor dem Komprimierungsvorgang ereignen, während alle Filteramplitudenwerte verfügbar sind.

Statt die Abtastrate zu verlangsamen, speichert die bevorzugte Anordnung alle Filteramplitudenwerte für 10 bis 15 Abtastungen bzw. Proben pro Sekunde für eine Sprachprobe von ungefähr 10 bis 15 Sekunden vor diesem Korrektur- und Komprimierungsprozeß. Wenn der Computerspeicherplatz kritischer als die Durchlaufgeschwindigkeit ist, sollten sich die Korrekturen und Komprimierung zwischen jedem Durchlauf ereignen, um den nächsten für einen großen starken Datenspeicher auszulöschen. Da die meisten kommerziell erhältlichen Minicomputer mit durchschnittlichem Preis einen ausreichenden Speicher aufweisen, speichert die bevorzugte und hierin geoffenbarte Anordnung alle Daten und be- bzw. verarbeitet nachher die Daten.

Die meisten Tierstimmensignale von Interesse, inklusive der menschlichen, beinhalten eine größte Amplitudenspitze, wahrscheinlich nicht an jedem Ende der Frequenzdomäne. Diese Spitze kann

durch irgendeinen einfachen und üblichen numerischen Sortieralgorithmus bestimmt werden, wie es in dieser Erfindung vorgenommen wird. Die für Amplitude und Frequenz repräsentativen Werte werden dann in der Zahl drei von sechs Speicherstellensätzen angeordnet, um die Amplituden und Frequenzen von sechs Spitzen bzw. Peaks zu halten.

Die höchste Frequenzspitze über 8 kHz wird in der Speicherstelle Nummer sechs angeordnet und als Hochfrequenzspitze gekennzeichnet. Die niedrigste Spitze wird in dem ersten Satz von Speicherstellen bzw. Speicherplätzen angeordnet. Die anderen drei werden aus Spitzen zwischen diesen ausgewählt. Dieser Kompressionsfunktion folgend wird das Stimmsignal durch einen für Amplitude und Frequenz repräsentativen Wert von jeder der sechs Spitzen plus einer Gesamtenergieamplitude aus dem ungefilterten Gesamtsignal für beispielsweise zehn mal pro Sekunde für eine Probe bzw. Abtastung von zehn Sekunden repräsentiert. Dies liefert eine Gesamtzahl von 1300 Werten.

Die Algorithmen erlauben Variationen in der Abtastlänge im Fall, daß der Bediener den Abtastlängenschalter mit dem Übersteuerungsausschalter übersteuert bzw. außer Kraft setzt, um eine Fortsetzung während einer unerwarteten Geräuschunterbrechung zu verhindern. Die Algorithmen erledigen dies durch Verwendung von Durchschnitten, welche nicht signifikant empfindlich auf Änderungen in der Abtastanzahl über vier oder fünf

Sekunden eines Schall- bzw. Klangsignals sind. Der Grund für eine größere Sprachabtastung, wenn möglich, ist, um den Durchschnitts-"Stil" der Sprache des Sprechers aufzufangen, welcher typischerweise innerhalb von 10 bis 15 Sekunden offensichtlich bzw. evident wird.

Die Ausgabe dieser Komprimierungsfunktion wird in die Elementanordnung und einen Speicheralgorithmus eingespeist, welcher zusammenbaut (a) vier Stimmqualitätswerte, welche unten zu beschreiben sind; (b) eine Klang- bzw. Schall-"Pause" oder ein Ein-zu-Aus-Verhältnis; (c) "Variabilität" – den Unterschied zwischen der Spitze jeder Amplitude für den gegenwärtigen Durchlauf und jenen des letzten Durchlaufs; Unterschiede zwischen der Frequenzzahl einer jeden Spitze für den gegenwärtigen Durchlauf und jener des letzten Durchlaufs; und Unterschied zwischen der ungefilterten Gesamtenergie des vorliegenden bzw. gegenwärtigen Durchlaufs und jener des letzten Durchlaufs; (d) eine "Silbenänderungsannäherung" durch ein Erhalten des Verhältnisses von Malen, daß sich die zweite Spitze mehr als 0,4 zwischen Durchläufen zu der Gesamtanzahl von Durchläufen mit Ton ändert; und (e) "Hochfrequenzanalyse" – das Verhältnis der Anzahl von Klang-Ein-Durchläufen, welche einen Nicht-Nullwert in dieser Spitze für die Spitzenamplitude Nummer sechs beinhalten. Dies ist eine Gesamtzahl von 20 Elementen, die pro Durchlauf verfügbar. Diese werden dann zum Dimensionszusammenbau-Algorithmus durchgeleitet.

Die vier Stimmqualitätswerte, welche als Elemente verwendet werden, sind (1) die "Ausbreitung bzw. Verbreitung" – der Abtastmittelwert von allen Unterschieden von Durchläufen zwischen ihrem Durchschnitt der eine Frequenz repräsentierenden Werte oberhalb der Maximalamplitudenspitze und dem Durchschnitt von jenen darunter, (2) die "Balance" – der Abtastmittelwert von allen Durchschnittsamplitudenwerte der Durchläufe von Spitzen 4, 5 & 6 geteilt durch den Durchschnitt der Spitzen 1 & 2. (3) "Hüllkurvenebenheit-hoch" – der Abtastmittelwert von allen Durchschnitten der Durchläufe von ihren Amplituden oberhalb der größten Spitze, geteilt durch die größte Spitze, (4) "Hüllkurven-ebenheit niedrig" – der Abtastmittelwert von allen Durchschnitten der Durchläufe von ihren Amplituden unterhalb der größten Spitze, geteilt durch die größte Spitze.

Die Stimmenstildimensionen werden "Resonanz" und "Qualität" bezeichnet und werden durch einen Algorithmus zusammengebaut, welcher eine Koeffizientenmatrix beinhaltet, die auf ausgewählten Elementen arbeitet.

Die "Sprachstil"-Dimensionen werden "Variabilität monoton", "abgehackt-sanft", "Stakkato halten", "Anstieg sanft", "Affektivitätssteuerung" bezeichnet. Diese fünf Dimensionen, wobei Namen zu jedem Ende von jeder Dimension gehören, werden gemessen und durch einen Algorithmus zusammengebaut, welcher eine Koeffizientenmatrix involviert, die auf 15 von den 20 Klangelementen arbeitet, die in Tabelle 6 und dem Spezifikationsabschnitt detailliert sind.

Die Wahrnehmungsstildimensionen werden "Eko-Struktur", "invariante Empfindlichkeit", "anders selbst", "sensorisch-intern", "Haß-Liebe", "Unabhängigkeit-Abhängigkeit" und "emotional-physisch" bezeichnet. Diese sieben Wahrnehmungsdimensionen mit Namen, die sich auf die Endbereiche der Dimensionen beziehen, werden gemessen und durch einen Algorithmus zusammengebaut bzw. zusammengestellt, welcher eine Koeffizientenmatrix involviert und auf ausgewählten Klangelementen von Stimme und Sprache (detailliert in Tabelle 7) und dem Spezifikationsabschnitt arbeitet.

Eine kommerziell erhältliche, typische Computertastatur oder Folientastatur gestattet es dem Anwender der vorliegenden Offenbarung, jeden und alle Koeffizienten für eine Neudefinition von irgendeiner zusammengesetzten Sprach-, Stimmen- oder Wahrnehmungsdimension für Forschungszwecke abzuändern. Auswahlschalter gestatten es, jedes oder alle Elemente oder Dimensionswerte für eine Stimmprobe eines vorgegebenen Subjekts anzuzeigen. Der digitale Prozessor regelt bzw. steuert die Analog-Zu-Digital-Wandlung des Klangsignals und regelt bzw. steuert auch den Wiederzusammenbau bzw. Neuzusammenbau der Stimmklangelemente in numerische Werte der Stimmen-, Sprach- und Wahrnehmungsdimensionen.

Der Mikrocomputer koordiniert auch die Tastatureingaben des Bedieners bzw. Betätigers und die gewählte Ausgabeanzeige von Werten, und Koeffizientenmatrixauswahl, um mit den Algorithmen zusammenzuwirken, welche die Stimmen-, Sprach- und Wahrnehmungsdimensionen zusammensetzen. Der Ausgabeauswahlschalter richtet einfach die Ausgabe zu jedem oder allen Ausgabesteckern, welche geeignet sind, um das Signal zu typischen kommerziell erhältlichen Monitoren, Modems, Druckern oder vorgegeben zu einer lichtausgebenden bordeigenen Anordnung zu richten.

Durch ein Entwickeln von Gruppenprofilstandards unter Verwendung dieser Erfindung kann ein Forscher Ergebnisse in Veröffentlichungen durch Berufe bzw. Beschäftigungen, Fehlfunktionen, Aufgabenstellungen, Hobbyinteressen, Kulturen, Sprachen, Geschlecht, Alter, Tierart, usw. auflisten. Oder der (die) Anwender (in) kann seine/ihre Werte mit jenen vergleichen, welche durch andere veröffentlicht wurden, oder mit jenen, welche in die Maschine eingebaut sind. von wenigstens einem Merkmal der Stimmsignale, wie beispielsweise in der oben besprochenen n

Bezugnehmend nun auf 12 der Zeichnungen wird eine stimmliche Äußerung in den Stimmklanganalysator durch ein Mikrofon 1210 eingebracht, und durch einen Mikrofonverstärker 1211 zur Signalverstärkung, oder von einer aufgezeichneten Eingabe durch einen Bandeingabestecker 1212 zur Verwendung einer vorab gezeichneten stimmlichen Äußerung eingegeben. Eine Eingabepegelregelung bzw. -steuerung 1213 stellt den Stimmsignalpegel auf den Filtertreiberverstärker 1214 ein. Der Filtertreiberverstärker 1214 verstärkt das Signal und legt das Signal an das V. U.-Meter 1215 zum Messen des korrekten Betriebssignalpegels an.

Die Durchlaufrate pro Sekunde und die Anzahl von Durchläufen pro Abtastung bzw. Probe wird durch den Betätiger bzw. Bediener mit den Durchlaufraten- und Abtastzeitschalter 1216 geregelt bzw. gesteuert. Der Bediener startet eine Abtastung mit dem Abtaststartschalter und der Stopübersteuerung 1217. Das Übersteuerungsmerkmal gestattet es dem Bediener, manuell die eingestellte Abtastzeit zu übersteuern, und die Abtastung bzw. Probennahme zu stoppen, um ein Kontaminieren bzw. Verunreinigen einer Abtastung bzw. Probe mit unerwarteten Klangstörungen zu

verhindern, beinhaltend gleichzeitige Lautsprecher. Dieser Schalter verbindet und unterbricht auch die Stromversorgung des Mikroprozessors mit 110 Volt elektrischen Standardeingabestiften bzw. -anschlüssen.

Die Ausgabe des Filtertreiberverstärkers 1214 wird auch an eine kommerziell erhältliche, durch einen Mikroprozessor gesteuerte Filterbank und einen Digitalisierer 1218 angewendet bzw. angelegt, welche(r) das elektrische Signal in 1/3-Oktavenbereiche über den hörbaren Frequenzbereich für den Organismen segmentiert bzw. unterteilt, welcher abgetastet wird, und die Spannungsausgabe von jedem Filter digitalisiert. In einem spezifischen Arbeitssystem reichen 25 1/3-Oktavenfilter eines Eventide-Spektrumanalysators mit Filtermittelfrequenzen von 63 Hz bis 16.000 Hz. Weiters wurde ein AKAI-Mikrofon und Bandaufzeichnungsgerät mit eingebautem Verstärker als die Eingabe in die Filterbank und den Digitalisierer 1218 verwendet. Die Anzahl von Durchläufen pro Sekunde, welche die Filterbank verwendet, beträgt ungefähr zehn Durchläufe pro Sekunde. Andere mikroprozessorgesteuerte Filterbanken und Digitalisierer können bei unterschiedlichen Geschwindigkeiten arbeiten.

Irgendein von verschiedenen kommerziell erhältlichen Mikroprozessoren ist geeignet, um die oben angeführte Filterbank und den Digitalisierer zu regeln bzw. zu steuern.

Wie bei jedem komplexen Klang bzw. Schall wird eine Amplitude über den hörbaren Frequenzbereich für einen "Zeitanteil bzw. Zeitstück" von 0,1 einer Sekunde nicht konstant oder flach sein, sondern wird eher Spitzen und Täler sein bzw. aufweisen. Die für eine Frequenz repräsentativen Werte der Spitzen dieses Signals, 1219, werden genauer gemacht, indem die Amplitudenwerte auf jeder Seite der Spitzen festgestellt werden, und die Spitzenwerte zu dem benachbarten Filterwert eingestellt werden, welcher die größere Amplitude aufweist. Dies wird gemacht, da, wie es für benachbarte 1/3-Oktavenfilter charakteristisch ist, Energie bei einer gegebenen Frequenz in einem gewissen Ausmaß in benachbarte Filter abhängig von den Abschneide- bzw. Wegschneidequalitäten der Filter überläuft. Um diesen Effekt zu minimieren, wird von der Frequenz eines Spitzenfilters angenommen, die Mittenfrequenz zu sein, nur wenn die zwei benachbarten Filter Amplituden innerhalb von 10% von ihrem Durchschnitt aufweisen. Um diskrete, gleich beabstandete, kleine Werte für ein Linearisieren und Normalisieren der Werte zu garantieren, die die ungleichen Frequenzintervalle repräsentieren, werden jedem der 25 Filter Zahlen- bzw. Nummernwerte 1 bis 25 gegeben und diese Nummern werden für den Rest der Bearbeitung verwendet. Auf diese Weise wird der 3.500 Hz-Unterschied zwischen den Filtern 24 und 25 ein Wert von 1, welcher wiederum ebenfalls gleich der 17 Hz-Differenz zwischen dem ersten und zweiten Filter ist.

Und mehr als fünf Sub-Unterteilungen von jeder Filteranzahl zu verhindern und damit fortzufahren, gleiche bewertete Schritte bzw. Stufen zwischen jeder Sub- bzw. Unterabteilung der 1 bis 25 Filterzahlen aufrechtzuerhalten, werden diese in 0,2 Schritte unterteilt und weiter zugeordnet wie folgt. Wenn die Amplitudendifferenz der zwei benachbarten Filter zu einem Spitzenfilter größer als angenommen, näher dem Punkt auf halbem Weg zur nächsten Filterzahl zu sein, als sie es von dem Spitzenfilter ist. Dies würde die

Filterzahl eines Spitzenfilters, sagen wir, Filternummer 6,0, dazu veranlassen, auf 6,4 erhöht oder auf 5,6 vermindert zu werden, wenn das größere benachbarte Filter eine höhere bzw. niedrigere Frequenz repräsentiert. Allen anderen Filterwerten von Spitzenfiltern wird automatisch der Wert seiner Filternummer +0,2 und -0,2 gegeben, wenn die größere der benachbarten Filteramplituden jeweils eine höhere oder niedrigere Frequenz repräsentiert.

Das segmentierte und digital repräsentierte bzw. dargestellte Stimmäußerungssignal 1219 wird nach der oben angeführten Frequenzkorrektur 1220 komprimiert, um Speicherplatz zu sparen, indem alle

außer sechs Amplitudenspitzen verworfen werden. Der Erfinder fand heraus, daß sechs Spitzen ausreichend waren, um die Stilcharakteristika aufzufangen, solange die folgenden Charakteristika bzw. Merkmale beobachtet werden. Wenigstens eine Spitze ist nahe der fundamentalen bzw. Grundfrequenz; genau eine Spitze ist zwischen dem Bereich der Grundfrequenz und der Spitzenamplitudenfrequenz gestattet, wo die nächste zur maximalen bzw. Maximalspitze konserviert wird; und die ersten zwei Spitzen oberhalb der Maximalspitze werden gespeichert plus die Spitze am nächsten zu dem 16.000 Hz-Ende oder dem 25. Filter, wenn oberhalb von 8 kHz, für eine Gesamtanzahl von sechs Spitzen, die im Mikroprozessorspeicher aufbewahrt und gespeichert werden. Dies wird garantieren, daß die Maximalspitze immer die dritte Spitze ist, welche im Speicher gespeichert wird, und daß die sechste gespeicherte Spitze für eine Hochfrequenzanalyse verwendet werden kann, und die die erste die niedrigste und nächste zur Grundlegenden bzw. Fundamentalen ist.

Nachfolgend auf die Kompression des Signals, um einen Amplitudenwert eines vollständigen Bands, die Filterzahl und Amplitudenwert von sechs Spitzen, und jeden dieser dreizehn Werte für 10 Abtastungen für eine 10 Sekunden-Abtastung (1300 Werte), 1221 von 12, zu beinhalten, beginnt der Klangelementzusammenbau.

Um bei Stimmstil-"Qualitäts"-Elementen anzukommen, verwendet dieses System Beziehungen zwischen dem niedrigen Satz und höheren Satz von Frequenzen in der stimmlichen Äußerung. Die Sprachstilelemente werden andererseits durch eine Kombination von Messungen bezüglich des Musters von Stimmenergieauftritten, wie beispielsweise Pausen und Abklingraten bestimmt. Diese Stimmstil-"Qualitäts"-Elemente tauchen aus der Spektrumanalyse, 13, 1330, 1331, und 1332 auf. Die Sprachstilelemente tauchen aus anderen vier Analysenfunktionen auf, wie dies in 12, 1233, 1234, 1235, und 1236 und Tabelle 6 gezeigt ist.

Die gespeicherten Stimmstilqualitätsanalyse-Elemente werden bezeichnet und abgeleitet als: (1) die Spektrum-"Verteilung" – der Abtastmittelwert des Abstands in Filterzahlen zwischen dem Durchschnitt der Spitzenfilteranzahlen oberhalb und dem Durchschnitt der Spitzenfilterzahlen unter der maximalen Spitze, für jeden Durchlauf, 13, 1330; (2) die Energie-"Balance" des Spektrums – der Mittelwert für eine Abtastung aller Verhältnisse des Durchlaufs der Summe der Amplituden von jenen Spitzen oberhalb der Summe der Amplituden unterhalb der Maximalspitze, 1331; (3) die Spektrumhüllkurve "Flachheit" – der arithmetische Mittelwert für jeden von zwei Sätzen von Verhältnissen für jede Abtastung bzw. Probe – die Verhältnisse der Durchschnittsamplitude von diesen Spitzen oberhalb (hoch) zur Maximalspitze, und von jenen unterhalb (niedrig) der Maximalspitze zur Maximalspitze für jeden Durchlauf, 1332.

Die Sprachstilelemente, welche gespeichert sind, werden jeweils bezeichnet und abgeleitet: (1) Spektrumvariabilität – die sechs Mittelwerte einer Äußerungsabtastung bzw. -probe, der numerischen Differenzen bzw. Unterschiede zwischen jeder Filterzahl einer Spitze, bei einem Durchlauf, zu jeder entsprechenden Filternummer bzw. -zahl einer Spitze beim nächsten Durchlauf, und auch die sechs Amplitudenwertdifferenzen für diese sechs Spitzen und auch beinhaltend die vollständigen Spektrumamplitudendifferenzen für jeden Durchlauf, um eine Abtastsumme von 13 Mittelwerten, 1333 zu erzeugen; (2) Äußerungspausenverhältnis-Analyse – das Verhältnis der Anzahl von Durchläufen in der Abtastung, bei welcher die vollständigen Energieamplitudenwerte Pausen waren (unter zwei Einheiten des Amplitudenwerts) zu der Zahl, welche Schallenergie aufwies (größer als eine Einheit des Werts), 1334; (3) Silbenänderungsannäherung – das Verhältnis der Anzahl von Durchläufen, bei welchen die dritte Spitze den Nummernwert um mehr als 0,4 zur Durchläufen geändert hat, welche Schall bzw. Klang während der Abtastung aufwies, 1335; (4) und, Hochfrequenzanalyse – das Verhältnis der Anzahl von Durchläufen für die Abtastung, bei welchem die sechste Spitze einen Amplitudenwert zur Gesamtanzahl von Durchläufen aufwies, 1336.

Klangstile werden in die sieben Dimensionen unterteilt, wie dies in Tabelle 6 dargestellt ist. Von diesen wurde bestimmt, die am empfindlichsten für einen assoziierten Satz von sieben Wahrnehmungs- oder Erkennungsstildimensionen zu sein, welche in Tabelle 7 aufgelistet sind.

Das Verfahren zum Beziehen der Klangstilelemente auf Stimmen-, Sprach- und Wahrnehmungsdimension zur Ausgabe, 12, 1228, erfolgt durch Gleichungen, welche jede Dimension als eine Funktion ausgewählter Klangstilelemente bestimmen, 13, 1330 bis 1336. Tabelle 6 bezieht die Sprachstilelemente, 1333 bis 1336 von 13, auf die Sprachstildimensionen.

Tabelle 7 stellt die Beziehung zwischen sieben Wahrnehmungsstildimensionen und den Klangstilelementen, 1330 bis 1336 dar. Erneut ist der Zweck, eine optionale Eingabekoeffizientenanzahl zu haben, welche Nullen beinhaltet, es dem Vorrichtungsbediener zu erlauben, Änderungen in diesen Koeffizienten zu Forschungszwecken, 1222, 1223 zu schalten oder einzugeben. Der schlaue Bediener kann unterschiedliche Wahrnehmungsdimensionen oder sogar Persönlichkeits- oder Erkenntnisdimensionen, oder Faktoren (wenn er diese Ausdrucksweise bevorzugt) entwickeln, welche alle gemeinsam unterschiedliche Koeffizienten erfordern. Dies wird vorgenommen bzw. durchgeführt, indem der gewünschte Satz von Koeffizienten eingegeben wird und notiert wird, auf welche Dimension (1226) er diese bezieht. Beispielsweise muß die Dimension anderer – selbst von Tabelle 7 nicht eine gewünschte Dimension durch einen Forscher sein, welcher diese durch eine Anwenderwahrnehmungsdimension ersetzen möchte, welche er introvertiert – extrovertiert nennt. Durch ein Ersetzen des Koeffizientensatzes für den Satz anderer – selbst, durch Versuchssätze, bis eine annehmbar hohe Korrelation zwischen der ausgewählten Kombination von gewichteten Klangstilelementen und seiner extern bestimmten Dimension introvertiert-extrovertiert besteht, kann der Forscher somit diesen Slot für die neue Dimension introvertiert – extrovertiert verwenden, indem er diese effektiv neu benennt. Dies kann in dem Ausmaß vorgenommen werden, daß der Satz von Klangelementen dieser Erfindung empfindlich ist für eine Anwenderdimension von introvertiert-extrovertiert, und der Koeffizientensatz des Forschers reflektiert die geeignete Beziehung. Dies wird möglich sein mit ziemlich vielen, durch einen Anwender bestimmten Dimensionen in einem nützlichen Ausmaß, wodurch es dem System ermöglicht wird, produktiv in einer Forschungsumgebung zu funktionieren, wo neue Wahrnehmungsdimensionen, bezogen auf Klangstilelemente, erforscht, entwickelt oder bewertet werden.

Tabelle 6 * ##STR1##

* DS1 = Variabilität monoton

* DS2 = abgehackt sanft bzw. glatt

* DS3 = Stakkato aufrechterhalten

* DS4 = Anstieg sanft

* DS5 = Affektivitätsregelung bzw. -steuerung

* (2) Nr. 1 bis 6 = Spitzenfilterunterschiede 1–6, und Amp1 bis 6 = Spitzenamplitudendifferenzen bzw. -unterschiede 1–6.

* Amp7 = Volle Bandpaßamplitudendifferenzen.

Tabelle 7 * ##STR2##

- * DP1 = EcoStruktur hoch-niedrig;
- * DP2 = Invariantempfindlichkeit hoch-niedrig;
- * DP3 = anders-selbst;
- * DP4 = sensorisch-intern;
- * DP5 = Haß-Liebe;
- * DP6 Abhängigkeit-Unabhängigkeit;
- * DP7 = emotionell-physisch. Anzahl von
- * (2) Nr. 1 bis 6 = Spitzenfilterdifferenzen 1–6; Amp1 bis 6 = Spitzenamplitudendifferenzen 1–6; und Amp7 vollständige Bandpaßamplitudendifferenzen.

Die für den Anwender dieses Systems verfügbaren primären Resultate sind die Dimensionswerte, 1226, welche selektiv durch einen Schalter, 1227 verfügbar sind, um auf einer Standardlichtanzeige und auch selektiv für Monitor, Drucker, Modem und andere Standardausgabevorrichtungen, 1228 angezeigt zu werden. Diese können verwendet werden, um zu bestimmen, wie nahe die Stimme des Subjekts zu einer beliebigen oder allen der Klang- oder Wahrnehmungsdimensionen von den eingebauten oder veröffentlichten oder persönlich entwickelten Regelungen bzw. Steuerungen oder Standards ist, welche dann verwendet werden können, um bei einer Verbesserung einer Emotionserkennung zu helfen.

In einer anderen beispielhaften Anordnung werden Biosignale, die von einem Anwender empfangen werden, verwendet, um dabei zu helfen, Emotionen in der Sprache des Anwenders zu bestimmen. Die Erkennungsrate eines Spracherkennungssystems wird durch ein Kompensieren von Änderungen in der Sprache des Anwenders verbessert, welche aus Faktoren, wie beispielsweise Emotion, Angst oder Ermüdung bzw. Müdigkeit resultieren. Ein von einer Äußerung eines Anwenders abgeleitetes Sprachsignal wird durch einen Vorprozessor modifiziert und einem Spracherkennungssystem bereitgestellt, um die Erkennungsrate zu verbessern. Das Sprachsignal wird basierend auf einem Biosignal modifiziert, welches für den emotionalen Zustand des Anwenders anzeigend bzw. hinweisend ist.

Detaillierter illustriert 14 ein Spracherkennungssystem, wo Sprachsignale vom Mikrophon 1418 und Biosignale vom Biomonitor 1430 durch einen Vorprozessor 1432 empfangen bzw. aufgenommen werden. Das Signal vom Biomonitor 1430 zum Vorprozessor 1432 ist ein Biosignal, welches für die Impedanz zwischen zwei Punkten auf der Oberfläche der Haut eines Anwenders anzeigend ist. Der Biomotor 1430 mißt die Impedanz unter Verwendung eines Kontakts 1436, welcher an einem der Finger des Anwenders befestigt ist, und eines Kontakts 1438, welche an einem anderen Finger des Anwenders befestigt ist. Ein Biomonitor, wie beispielsweise ein Biofeedbackmonitor bzw. Biorückkopplungsmonitor, welcher durch Radio Shack, welche eine Abteilung von Tandy Corporation ist, unter dem Handelsnamen (MICRONATA.RTM. BIOFEEDBACK MONITOR) Modell Nummer 63-664 verkauft wird, kann verwendet werden. Es ist auch möglich, die Kontakte an anderen Positionen auf der Haut des Anwenders festzulegen bzw. zu befestigen. Wenn der Anwender aufgeregt oder ängstlich wird, nimmt die Impedanz zwischen den Punkten 1436 und 1438 ab und die Abnahme wird durch den Monitor 1430 detektiert, welcher ein Biosignal erzeugt, welches anzeigend für bzw. hinweisend auf eine verringerte Impedanz ist. Der Vorprozessor 1432 verwendet das Biosignal von dem Biomotor 1430, um das vom Mikrophon 1418 empfangene Sprachsignal zu modifizieren, wobei das Sprachsignal modifiziert wird, um die

Änderungen in der Sprache des Anwenders aufgrund der aus Faktoren, wie beispielsweise Müdigkeit oder einer Änderung des emotionalen Zustands resultierenden Änderungen zu kompensieren. Beispielsweise kann der Vorprozessor 1432 die Tonhöhe des Sprachsignals von dem Mikrophon 1418 absenken, wenn das Biosignal von dem Biomonitor 1430 anzeigt, daß der Anwender in einem aufgeregten Zustand ist, und der Vorprozessor 1432 kann die Tonhöhe des Sprachsignals vom Mikrophon 1418 erhöhen, wenn das Biosignal von dem Biomonitor 1430 anzeigt, daß der Anwender in einem weniger aufgeregten Zustand ist, wie beispielsweise bei Ermüdung. Der Vorprozessor 1432 stellt dann das modifizierte Sprachsignal der Audiokarte 1416 auf herkömmliche Weise zur Verfügung. Für Zwecke, wie beispielsweise Initialisierung oder Kalibrierung, kann der Vorprozessor 1432 mit dem PC 1410 unter Verwendung eines Interface bzw. einer Schnittstelle, wie beispielsweise einer RS232-Schnittstelle, kommunizieren. Der Anwender 1434 kann mit dem Vorprozessor 1432 durch ein Beobachten der Anzeige 1412 und durch Eingeben von Befehlen unter Verwendung der Tastatur 1414 oder Folientastatur 1439 oder einer Maus kommunizieren.

Es ist auch möglich, das Biosignal zu verwenden, um das Sprachsignal durch Regeln bzw. Steuern des Gewinns bzw. der Verstärkung und/oder Frequenzantwort des Mikrophons 1418 vorzuverarbeiten. Der Gewinn oder die Verstärkung des Mikrophons kann in Antwort auf das Biosignal erhöht oder verringert werden. Das Biosignal kann auch verwendet werden, um die Frequenzantwort des Mikrophons zu ändern. Beispielsweise kann, wenn das Mikrophon 1418 ein ATM71, erhältlich von AUDIO-TECHNICA U.S., Inc., ist, das Biosignal verwendet werden, um zwischen einer relativ flachen Antwort und einer abgerollten Antwort zu schalten, wobei die abgerollte Antwort weniger Gewinn bzw. Verstärkung bei Sprachsignalen niedriger Frequenz zur Verfügung stellt.

Wenn der Biomonitor 1430 der oben genannte Monitor, erhältlich von Radio Shack, ist, ist das Biosignal in der Form einer Serie von rampenähnlichen Signalen, wobei jede Rampe ungefähr 0,2 ms andauert. 15 illustriert das Biosignal, wo eine Serie von rampenähnlichen Signalen 1542 durch eine Zeit T getrennt ist. Der Anteil bzw. das Ausmaß der Zeit T zwischen den Rampen 1542 bezieht sich auf die Impedanz zwischen den Punkten 1438 und 1436. Wenn der Anwender in einem aufgeregteren Zustand ist, wird die Impedanz zwischen den Punkten 1438 und 1436 verringert und die Zeit T wird verringert. Wenn sich der Anwender in einem weniger aufgeregten Zustand befindet, wird die Impedanz zwischen den Punkten 1438 und 1436 gesteigert und die Zeit T wird gesteigert bzw. vergrößert.

Die Form eines Biosignals von einem Biomonitor kann in anderen Formen als einer Serie von rampenähnlichen Signalen sein. Beispielsweise kann das Biosignal ein analoges Signal sein, welches in der Periodizität, Amplitude und/oder Frequenz basierend auf Messungen variiert, welche durch den Biomonitor vorgenommen wurden, oder es kann ein digitaler Wert basierend auf durch den Biomonitor gemessenen Bedingungen sein.

Der Biomonitor 1430 beinhaltet den Schaltkreis von 16, welcher das Biosignal erzeugt, das die Impedanz zwischen den Punkten 1438 und 1436 anzeigt. Der Schaltkreis besteht aus zwei Abschnitten. Der erste Abschnitt wird verwendet, um die Impedanz zwischen den Kontakten 1438 und 1436 zu erfassen, und der zweite Abschnitt wirkt als ein Oszillator, um eine Serie von Rampensignalen am Ausgangs- bzw. Ausgabeverbinder 1648 zu erzeugen, wo die Frequenz der Oszillation durch den ersten Abschnitt geregelt bzw. gesteuert wird.

Der erste Abschnitt regelt bzw. steuert den Kollektorstrom $I_{c,Q1}$ und die Spannung $V_{c,Q1}$ des Transistors Q1 basierend auf der Impedanz zwischen den Kontakten 1438 und 1436. In dieser Ausführungsform besteht der Impedanzsensor 1650 einfach aus Kontakten 1438 und 1436, welche auf der Haut des Sprechers positioniert sind bzw. werden. Da sich die Impedanz zwischen den

Kontakten 1438 und 1436 relativ langsam im Vergleich zu der Oszillations- bzw. Schwingungsfrequenz des Abschnitts 2 ändert, sind der Kollektorstrom $I_{c,Q1}$ und die Spannung $V_{c,Q1}$ virtuell bzw. nahezu konstant, soweit der Abschnitt 2 betroffen ist. Der Kondensator C3 stabilisiert weiter diese Ströme und Spannungen.

Der Abschnitt 2 wirkt als ein Oszillator. Die reaktiven Komponenten, L1 und C1, schalten den Transistor Q3 ein und aus, um eine Oszillation bzw. Schwingung zu erzeugen. Wenn die Energie bzw. Leistung zuerst eingeschaltet wird, schaltet $I_{c,Q1}$ Q2 durch Ziehen von Basisstrom $I_{b,Q2}$ ein. In ähnlicher Weise schaltet $I_{c,Q2}$ den Transistor Q3 ein, indem ein Basisstrom $I_{b,Q3}$ zur Verfügung gestellt wird. Anfänglich gibt es keinen Strom durch den Induktor bzw. die Induktivität L1. Wenn Q3 eingeschaltet ist, wird die Spannung VCC geringer als eine kleine gesättigte bzw. Sättigungs-Transistorspannung $V_{c,Q3}$ über L1 angelegt bzw. angewendet. Als ein Ergebnis steigt der Strom I_{L1} in Übereinstimmung mit

Da bzw. wenn der Strom I_{L1} ansteigt bzw. zunimmt, nimmt der Strom I_{c1} durch den Kondensator C1 zu. Ein Steigern des Stroms I_{c1} reduziert den Basisstrom $I_{b,Q2}$ vom Transistor Q2, da der Strom $I_{c,Q1}$ virtuell bzw. nahezu konstant ist. Dies reduziert wiederum die Ströme $I_{c,Q2}$, $I_{b,Q3}$ und $I_{c,Q3}$. Als ein Ergebnis tritt mehr vom Strom I_{L1} durch den Kondensator C1 durch und reduziert weiter den Strom $I_{c,Q3}$. Diese Rückkopplung veranlaßt den Transistor Q3 dazu, abgeschaltet zu werden. Schließlich ist der Kondensator C1 vollständig geladen und die Ströme I_{L1} und I_{c1} fallen auf null, und erlauben es dadurch dem Strom $I_{c,Q1}$ erneut, den Basisstrom $I_{b,Q2}$ zu ziehen und die Transistoren Q2 und Q3 einzuschalten, was den Oszillationszyklus erneut startet.

Der Strom $I_{c,Q1}$, welcher von der Impedanz zwischen den Kontakten 1438 und 1436 abhängt, regelt bzw. steuert die Frequenz des Lastverhältnisses bzw. Arbeitszyklus des Ausgabesignals. Wenn die Impedanz zwischen den Punkten 1438 und 1436 abnimmt, nimmt die Zeit T zwischen den Rampensignalen ab, und wenn die Impedanz zwischen den Punkten 1438 und 1436 zunimmt, nimmt die Zeit T zwischen den Rampensignalen zu.

Der Schaltkreis wird durch eine Drei-Volt-Batteriequelle 1662 angetrieben, welche mit dem Schaltkreis über den Schalter 1664 verbunden ist. Ebenfalls beinhaltet ist ein variabler Widerstand 1666, welcher verwendet wird, um einen Arbeitspunkt für den Schaltkreis einzustellen. Es ist wünschenswert, den variablen Widerstand 1666 auf eine Position einzustellen, welche ungefähr in der Mitte seines Einstellbereichs ist. Der Schaltkreis schwankt dann von diesem Arbeitspunkt, wie früher beschrieben, basierend auf der Impedanz zwischen den Punkten 1438 und 1436. Der Schaltkreis beinhaltet auch einen Schalter 1668 und einen Lautsprecher 1670. Wenn ein zusammenpassender Verbinder nicht in den Verbinder 1648 eingesetzt ist, stellt der Schalter 1668 die Ausgabe des Schaltkreises am Lautsprecher 1670 eher als am Verbinder 1648 zur Verfügung.

17 ist ein Blockdiagramm des Vorprozessors 1432. Ein Analog-zu-Digital-Konverter bzw. -Wandler (A/D) 1780 empfängt ein Sprach- oder Äußerungssignal vom Mikrofon 1418, und ein Analog-zu-Digital-Wandler (A/D) 1782 empfängt ein Biosignal von dem Biomonitor 1430. Das Signal von dem A/D 1782 wird einem Mikroprozessor 1784 zur Verfügung gestellt. Der Mikroprozessor 1784 überwacht das Signal von dem A/D 1782, um zu bestimmen, welche Maßnahme durch die digitale Signalprozessor-Vorrichtung (DSP) 1786 vorgenommen werden sollte. Der Mikroprozessor 1784 verwendet einen Speicher 1788 für eine Programmspeicherung und für Zwischenregistervorgänge. Der Mikroprozessor 1784 kommuniziert mit dem PC 1410 unter Verwendung einer RS232-Schnittstelle. Die Software zum Regeln bzw. Steuern der Schnittstelle zwischen dem PC 1410 und dem Mikroprozessor 1784 kann auf dem PC 1410 in einer Mehrfachanwendungsumgebung unter Verwendung eines Softwarepakets ablaufen, wie beispielsweise einem Programm, das unter dem Handelsnamen (WINDOWS) durch Microsoft Corporation verkauft wird. Die Ausgabe von dem DSP 1786 wird zu einem analogen Signal durch einen Digital-zu-Analog-Wandler 1790

rückgewandelt. Nachdem der DSP 1786 das Signal von dem A/D 1780 modifiziert, wie durch den Mikroprozessor 1784 befohlen wurde, wird die Ausgabe des D/A-Wandlers 1790 zu der Audiokarte 1416

gesandt. Der Mikroprozessor 1784 kann einer von weithin verfügbaren Mikroprozessoren sein, wie beispielsweise die Mikroprozessoren, die von Intel Corporation erhältlich sind, und der DSP 1786 kann einer der weit verbreitet erhältlichen digitalen Signalverarbeitungschips sein, welche von Gesellschaften, wie beispielsweise TMS320CXX-Serie von Geräten von Texas Instruments sind.

Es ist möglich, um den Biomonitor 1430 und Vorprozessor 1432 auf einer einzelnen Karte zu positionieren, welche in einen leeren Kartensteckplatz im PC 1410 eingesetzt wird. Es ist auch möglich, die Funktionen des Mikroprozessors 1784 und des digitalen Signalprozessors 1786 unter Verwendung des PC 1410 statt durch spezialisierte Hardware durchzuführen.

Der Mikroprozessor 1784 überwacht das Biosignal von dem A/D 1782, um zu bestimmen, welche Maßnahme durch den DSP 1786 vorgenommen werden sollte. Wenn das Signal von dem A/D 1782 anzeigt, daß der Anwender in einem aufgeregteren Zustand ist, zeigt der Mikroprozessor 1784 dem DSP 1786 an, daß er das Signal von dem A/D 1780 verarbeiten sollte, so daß die Tonhöhe des Sprachsignals verringert wird. Wenn das Biosignal von dem A/D 1782 anzeigt, daß der Anwender in einem weniger aufgeregten oder ermüdeten Zustand ist, instruiert der Mikroprozessor 1784 den DSP 1786, die Tonhöhe des Sprachsignals zu erhöhen.

Der DSP 1786 modifiziert die Tonhöhe des Sprachsignals durch ein Erzeugen eines Sprachmodells. Der DSP verwendet dann das Modell, um das Sprachsignal mit einer modifizierten Tonhöhe wieder herzustellen. Das Sprachmodell wird unter Verwendung einer der linearen voraussagenden Codiertechniken erzeugt, welche in der Technik gut bekannt sind. Eine derartige Technik ist in einem Anwendungsbuch von Analog Device, Inc., geoffenbart, mit dem Titel "Digitale Signalverarbeitungsanwendungen unter Verwendung der ADSP 2100 Familie", Seiten 355–372, veröffentlicht durch Prentice-Hall, Englewood Cliffs, N. J., 1992. Diese Technik involviert ein Modellieren des Sprachsignals als ein FIR-Filter (finite bzw. endliche Impulsantwort) mit zeitveränderlichen Koeffizienten, wobei das Filter durch einen Zug von Impulsen erregt wird. Dann ist die Zeit T zwischen den Impulsen ein Maß der Tonhöhe oder Grundfrequenz. Die zeitveränderlichen Koeffizienten können unter Verwendung einer Technik berechnet werden, wie beispielsweise die Levinson-Durbin-Rekursion, welche in der oben angeführten Veröffentlichung von Analog Device, Inc. geoffenbart ist. Eine Zeit T zwischen den Impulsen, welche den Zug von Impulsen bilden, welche das Filter erregen, kann unter Verwendung eines Algorithmus berechnet werden, wie beispielsweise der SIFT-Algorithmus (vereinfachtes inverses Filternachführen) von John D. Markel, welcher geoffenbart ist in "Der SIFT-Algorithmus zur Grundfrequenzabschätzung" von John D. Markel, IEEE Transactions on Audio und Electroacoustics, Vol. AU-20, Nr. 5, Dezember 1972. Der DSP 1786 modifiziert die Tonhöhe oder Grundfrequenz des Sprachsignals durch ein Ändern der Zeit T zwischen den Impulsen, wenn er das FIR-Filter erregt, um das Sprachsignal wiederherzustellen. Beispielsweise kann die Tonhöhe um 1% erhöht werden, indem die Zeit T zwischen den Impulsen um 1% verringert wird.

Es sollte beachtet werden, daß das Sprachsignal auf andere Weisen als Änderungen in der Tonhöhe modifiziert sein kann. Beispielsweise können Tonhöhe, Amplitude, Frequenz und/oder Signalspektrum modifiziert sein. Ein Abschnitt des Signalspektrums oder des Gesamtspektrums kann abgeschwächt oder verstärkt sein.

Es ist auch möglich, andere Biosignale als ein Signal zu überwachen, welches für die Impedanz zwischen zwei Punkten auf einer Haut des Anwenders hinweisend bzw. anzeigend ist. Signale, die für eine autonome Aktivität anzeigend sind, können als Biosignale verwendet werden. Signale,

welche für eine autonome Aktivität anzeigend sind, wie beispielsweise Blutdruck, Pulszahl, Hirnwellen- oder andere elektrische Aktivität, Pupillengröße, Hauttemperatur, Transparenz oder Reflexionsvermögen einer bestimmten elektromagnetischen Wellenlänge, oder andere Signale, die für den emotionalen Zustand des Anwenders anzeigend sind, können verwendet werden.

18 illustriert Tonhöhenmodifikationskurven, welche der Mikroprozessor 1784 verwendet, um den DSP 1786 zu instruieren, die Tonhöhe des Sprachsignals basierend auf der Zeitperiode T, welche mit dem Biosignal assoziiert ist, zu ändern. Die horizontale Achse 1802 zeigt die Zeitperiode T zwischen den Rampen 1442 des Biosignals an und die vertikale Achse 1804 zeigt die Prozentänderung in der Tonhöhe an, welche durch den DSP 1786 eingebracht wird.

19 illustriert ein Flußdiagramm der Befehle, welche durch den Mikroprozessor 1784 ausgeführt werden, um eine in 18 illustrierte Betriebslinie bzw. -kurve zu errichten. Nach einer Initialisierung wird Schritt 1930 ausgeführt, um eine Linie zu errichten, die co-linear mit der Achse 1802 ist. Diese Linie zeigt an, daß eine Null-Tonhöhenänderung für alle Werte von T von dem Biosignal eingebracht ist. Nach dem Schritt 1930 wird ein Entscheidungsschritt 1932 ausgeführt, wo der Mikroprozessor 1784 bestimmt, ob ein Modifikationskommando bzw. -befehl von der Tastatur 1414 oder der Folientastatur 1439 empfangen wurde. Wenn kein Modifikationsbefehl empfangen wurde, wartet der Mikroprozessor 1784 in einer Schleife auf einen Modifikationsbefehl. Wenn ein Modifikationskommando empfangen wird, wird ein Schritt 1934 ausgeführt, um den Wert von T = Trefl zu bestimmen, welcher verwendet werden wird, um einen neuen Referenz- bzw. Bezugspunkt Refl zu errichten. Der Wert Trefl ist gleich dem gegenwärtigen Wert von T, der aus dem Biosignal erhalten wurde. Beispielsweise kann Trefl gleich 0,6 ms sein. Nach einer Bestimmung des Werts Trefl führt der Mikroprozessor 1784 einen Schritt 1938 aus, welcher den Anwender auffordert, eine Äußerung zu tätigen, so daß eine Tonhöhenprobe in Schritt 1940 entnommen werden kann. Es ist wünschenswert, eine Tonhöhenprobe zu erhalten, da die Tonhöhenprobe als eine Basis für die prozentuellen Änderungen der Tonhöhe verwendet wird, die entlang der Achse 1804 angezeigt ist. In Schritt 1942 instruiert der Mikroprozessor 1784 den DSP 1786, die Tonhöhe des Sprachsignals um einen Betrag gleich der gegenwärtigen Tonhöhenänderung, die mit dem Punkt Refl assoziiert ist, plus einer Erhöhung von fünf Prozent zu erhöhen, jedoch können kleinere oder größere Erhöhungen bzw. Schrittgrößen verwendet werden. (An diesem Punkt ist die mit dem Punkt Refl assoziierte Tonhöhenänderung null. Siehe bzw. Wiederaufrufen von Schritt 1930.) In Schritt 1944 fordert der Mikroprozessor 1784 den Anwender auf, einen Erkennungstest auszuführen, indem verschiedene Kommandos bzw. Befehle zu dem Spracherkennungssystem gesprochen werden, um zu bestimmen, ob eine annehmbare Erkennungsrate erzielt wurde. Wenn der Anwender den Test beendet, kann der Anwender die Beendigung des Tests dem Mikroprozessor 1784 anzeigen, indem ein Befehl, wie beispielsweise "Ende", unter Verwendung der Tastatur 1414 oder Folientastatur 1439 eingegeben wird.

Nach einem Ausführen des Schritts 1944 führt der Mikroprozessor 1784 einen Schritt 1946 aus, in welchem er den DSP 1786 instruiert, die Tonhöhe des eingehenden Sprachsignals um die Tonhöhenänderung zu vermindern, welche mit dem Punkt Refl assoziiert ist, minus einer Verminderung von fünf Prozent; jedoch können kleinere oder größere Beträge bzw. Anteile verwendet werden. (Beachten Sie, daß die Tonhöhenänderung, welche mit dem Punkt Refl assoziiert ist, null als ein Ergebnis des Schritts 1930 ist). In Schritt 1948 fordert der Mikroprozessor 1784 den Anwender auf, einen anderen Spracherkennungstest durchzuführen und ein "Ende"-Kommando einzugeben, wenn der Test abgeschlossen ist. In Schritt 1950 fordert der Mikroprozessor 1784 den Anwender auf, für den ersten oder zweiten Test zu stimmen, um anzuzeigen, welcher Test eine bessere Erkennungsfähigkeit hatte. In Schritt 1952 werden die Resultate der Wahl des Anwenders verwendet, um zwischen den Schritten 1954 und 1956 auszuwählen. Wenn der Test 1 als der beste bewertet wurde, wird der Schritt 1956 ausgeführt und die neue, mit dem Punkt Refl assoziierte, prozentuelle bzw. Prozentänderung wird dem

vorhergehenden Wert des Punkts Ref1 plus fünf Prozent oder die Erhöhung gleichgesetzt, welche in Schritt 1942 verwendet wurde. Wenn der Test 2 als der beste bewertet ist, wird der Schritt 1954 ausgeführt und der neue, mit Ref1 assoziierte, Prozentänderungswert wird dem alten Wert von Ref1 minus fünf Prozent oder die Verminderung gleichgesetzt, welche in Schritt 1946 verwendet wurde. Ein Bestimmen einer prozentuellen bzw. Prozentänderung, die mit $T = Tref1$ assoziiert ist, errichtet einen neuen Referenzpunkt Ref1. Beispielsweise ist, wenn der Test 1 als der beste bewertet wurde, der Punkt Ref1 am Punkt 1858 in 18 angeordnet. Nach Errichten der Position des Punkts 1858, welcher der neu errichtete Ref1 ist, ist bzw. wird die Linie 1860 in Schritt 1962 errichtet. Die Linie 1860 ist die Anfangstonhöhen-Modifikationslinie, welche verwendet wird, um die Tonhöhenänderungen für unterschiedliche Werte von T aus dem Biosignal zu berechnen. Anfänglich kann dieser Linie eine Neigung bzw. ein Anstieg von beispielsweise plus fünf Prozent pro Millisekunde gegeben sein; jedoch können andere Neigungen bzw. Anstiege verwendet werden.

Nach Errichten bzw. Aufbauen dieser anfänglichen Modifikationslinie geht der Mikroprozessor 1784 in eine Warteschleife, wo Schritte 1964 und 1966 ausgeführt werden. In Schritt 1964 prüft der Mikroprozessor 1784 nach einem Modifikationskommando bzw. -befehl, und in Schritt 1966 prüft er nach einem Abschaltkommando. Wenn ein Modifizierungsbefehl in Schritt 1964 nicht empfangen wird, überprüft der Prozessor nach dem Abschaltkommando in Schritt 1966. Wenn ein Abschaltkommando nicht empfangen ist bzw. wird, kehrt der Mikroprozessor zu Schritt 1964 zurück, und wenn ein Abschaltkommando empfangen wird, führt der Mikroprozessor Schritt 1930 aus, welcher die Tonhöhenänderung mit null für alle Werte von T von dem Biosignal gleichsetzt. Der Prozessor bleibt in dieser Schleife zum Prüfen nach Modifizierungs- und Abschaltkommandos, bis der Anwender mit der Erkennungsrate unzufrieden wird, welche aus dem Vorverarbeiten des Sprachsignals unter Verwendung der Kurve 1860 resultiert.

Wenn in Schritt 1964 ein Modifizierungskommando empfangen wird, wird ein Schritt 1968 ausgeführt. In Schritt 1968 wird der Wert von T bestimmt, um zu überprüfen, ob der Wert von T gleich oder nahezu gleich dem Wert von $Tref1$ des Punkts Ref1 ist. Wenn der Wert von T mit Ref1 übereinstimmt, wird der Schritt 1942

ausgeführt. Wenn der Wert von T nicht mit Ref1 übereinstimmt, wird der Schritt 1970 ausgeführt. In Schritt 1970 wird der Wert von $Tref2$ für einen neuen Referenz- bzw. Bezugspunkt Ref2 errichtet. Zum Zweck eines illustrativen Beispiels werden wir annehmen, daß $Tref2 = 1,1$ ms. Unter Bezugnahme auf 18 errichtet dies den Punkt Ref2 als Punkt 1872 auf der Linie 1860. In Schritt 1974 instruiert der Mikroprozessor 1784 den DSP 1786, die Tonhöhenänderung, welche mit dem Punkt Ref2 assoziiert ist, um plus 2,5 Prozent (andere Prozentwerte können verwendet werden) zu erhöhen. (Andere Prozentwerte können verwendet werden). In Schritt 1976 wird der Anwender aufgefordert, einen Erkennungstest durchzuführen und das "Ende"-Kommando bzw. -befehl beim Abschluß einzugeben. In Schritt 1978 instruiert der Mikroprozessor 1784 den DSP 1786, die Tonhöhe des Sprachsignals um einen Betrag gleich einer Tonhöhenänderung zu vermindern, die mit Ref2 minus 2,5 Prozent assoziiert ist. In Schritt 1980 wird der Anwender erneut aufgefordert, einen Erkennungstest durchzuführen und einen "Ende"-Befehl bei Beendigung einzugeben. In Schritt 1982 wird der Anwender aufgefordert anzuzeigen, ob der erste oder zweite Test die erstrebenswertesten Resultate aufwies. In Schritt 1984 entscheidet der Mikroprozessor 1784, einen Schritt 1986 auszuführen, wenn Test 1 als der beste bewertet wurde, einen Schritt 1988, wenn Test 2 als der beste bewertet wurde. In Schritt 1986 stellt der Mikroprozessor 1784 die Prozentänderung, welche mit Punkt Ref2 assoziiert ist, auf den früheren Wert ein, welcher mit Ref2 plus 2,5 Prozent oder der Erhöhung assoziiert ist, welche in Schritt 1974 verwendet wurde. In Schritt 1988 wird die Prozentänderung, die mit Ref2 assoziiert ist, gleich dem früheren Wert eingestellt, der mit Ref2 minus 2,5 Prozent oder der Verringerung assoziiert ist, die in Schritt 1978 verwendet wurde. Nach Vollendung der Schritte 1986 oder 1988 wird ein Schritt 1990 ausgeführt. In Schritt 1990 ist eine neue Tonhöhenmodifizierungslinie errichtet. Die neue Linie verwendet den Punkt, der mit Ref1

assoziiert ist, und den neuen Punkt, der mit Ref2 assoziiert ist. Beispielsweise ist, wenn angenommen wird, daß der Anwender Test 1 in Schritt 1984 ausgewählt hat, der neue mit Ref2 assoziierte Punkt der Punkt 1892 von 18. Die neue Tonhöhenumwandlungslinie ist nun die Linie 1898, welche durch die Punkte 1892 und 1858 durchtritt. Nach Ausführen des Schritts 1990 kehrt der Mikroprozessor 1684 zu der mit den Schritten 1964 und 1966 assoziierten Schleifenfunktion zurück.

Es sollte beachtet werden, daß eine lineare Modifikationslinie verwendet wurde; jedoch ist es möglich, nicht-lineare Modifikationslinien zu verwenden. Dies kann vorgenommen werden, indem die Punkte 1858 und 196 verwendet werden, um einen Anstieg für eine Linie rechts des Punkts 1858 zu errichten, und indem ein anderer Referenz- bzw. Bezugspunkt links des Punkts 1858 verwendet wird, um eine Steigung für eine Linie zu errichten, die sich links des Punkts 1858 erstreckt. Es ist auch möglich, positive und negative Grenzen auf der maximalen, prozentuellen Tonhöhenänderung anzuordnen. Wenn die Tonhöhenmodifikationslinie sich diesen Grenzen nähert, können sich diese ihr asymptotisch annähern oder einfach abrupt am Kontaktpunkt mit der Grenze ändern.

Es ist auch möglich, eine festgelegte Modifikationskurve zu verwenden, wie beispielsweise Kurve 1800, und dann den variablen Widerstand 1666 einzustellen, bis eine annehmbare Erkennungsrate erzielt wird.

Sprach- bzw. Stimmbenachrichtigungssystem

20 stellt ein System dar, welches Sprachnachrichten basierend auf Emotionscharakteristika der Sprachnachrichten handhabt. In Vorgang 2000 wird eine Vielzahl von Sprachnachrichten, welche über ein Telekommunikationsnetzwerk übertragen werden, empfangen. In Vorgang 2002 werden die Sprachnachrichten auf einem Speichermedium, wie beispielsweise dem oben dargelegten Bandaufzeichnungsgerät oder einer Festplatte beispielsweise gespeichert. Eine mit den Sprach- bzw. Stimmsignalen der Sprachnachrichten assoziierte Emotion wird in Vorgang 2004 bestimmt. Die Emotion kann durch irgendeines der oben dargelegten Verfahren bestimmt werden.

Die Sprachnachrichten werden in Vorgang bzw. Funktion 2006 basierend auf der bestimmten Emotion organisiert. Beispielsweise Nachrichten, in welchen die Stimme negative Emotionen, z.B. Traurigkeit, Ärger oder Angst, anzeigt, können gemeinsam in einer Mailbox bzw. einem Briefkasten und/oder einer Datenbank gruppiert werden. Ein Zugriff auf die organisierten Sprachnachrichten ist in Vorgang 2008 gestattet.

Die Sprachnachrichten können einem Telefonanruf folgen. Optional können die Sprachnachrichten einer ähnlichen Emotion zusammen organisiert sein. Ebenfalls optional können die Sprachnachrichten in Echtzeit unmittelbar nach Erhalt über das Telekommunikationsnetzwerk organisiert sein. Vorzugsweise ist eine Weise, in welcher die Sprachnachrichten organisiert sind, identifiziert, um den Zugriff auf die organisierten Sprachnachrichten zu erleichtern. Ebenfalls vorzugsweise wird die Emotion durch ein Extrahieren von wenigstens einem Merkmal aus Sprachsignalen bestimmt, wie dies zuvor besprochen wurde.

In einer beispielhaften Anordnung werden Tonhöhen- und LPC-Parameter (und üblicherweise auch andere Erregungsinformation) zur Übertragung und/oder Speicherung codiert, und werden decodiert, um eine nahe Nachbildung der ursprünglichen Spracheingabe zur Verfügung zu stellen.

Das vorliegende System ist besonders auf lineare voraussagende Codierungssysteme (LPC) für ein Analysieren oder Codieren analoger Sprachsignale (und Verfahren hierfür) bezogen. In einer LPC-Modellierung wird allgemein jede Abtastung bzw. Probe in einer Serie von Abtastungen (im

vereinfachten Modell) als eine lineare Kombination von vorhergehenden Abtastungen modelliert, plus einer Erregungsfunktion: wobei u_k das LPC-Restsignal ist. Das heißt, u_k repräsentiert die verbleibende bzw. Restinformation in dem eingegebenen bzw. Eingabesprachsignal, welches nicht durch das LPC-Modell vorhergesagt ist. Es soll beachtet werden, daß nur N ältere bzw. frühere Signale zur Vorhersage verwendet werden. Die Modellreihenfolge (typischerweise etwa 10) kann erhöht werden, um eine bessere Voraussage zu ergeben, jedoch wird etwas Information immer in dem Restsignal u_k für jede normale Sprachmodellierungsanwendung verbleiben.

Innerhalb des allgemeinen Rahmenwerks der LPC-Modellierung können viele besondere Implementierungen einer Stimmanalyse ausgewählt werden. In vielen von diesen ist es notwendig, die Tonhöhe des Eingabesprachsignals zu bestimmen. D.h., zusätzlich zu den Formanten-Frequenzen, welche tatsächlich mit Resonanzen des Stimmtrakts übereinstimmen, beinhaltet die menschliche Stimme auch eine Tonhöhe, die durch den Sprecher moduliert wird, welche mit der Frequenz übereinstimmt, bei welcher der Kehlkopf den Luftstrom moduliert. D.h., die menschliche Stimme kann als eine Erregungsfunktion betrachtet werden, welche an ein akustisches passives Filter angewendet bzw. angelegt wird, und die Erregungsfunktion wird im allgemeinen in der LPC-Restfunktion erscheinen, während die Charakteristika bzw. Merkmale des passiven akustischen Filters (d.h. die Resonanzcharakteristika von Mund, Nasenhohlraum, Brustkorb, usw.) durch die LPC-Parameter geformt werden wird. Es sollte beachtet werden, daß während stimmloser Sprache die Erregungsfunktion nicht eine gut definierte Tonhöhe aufweist, sondern stattdessen als breitbandiges, weißes Rauschen oder rosa Rauschen modelliert ist.

Eine Abschätzung der Tonhöhenperiode ist nicht vollständig trivial. Unter den Problemen ist die Tatsache, daß die erste Formante häufig bei einer Frequenz nahe derjenigen der Tonhöhe auftreten wird. Aus diesem Grund wird die Tonhöhenabschätzung häufig an dem LPC-Restsignal durchgeführt, da der LPC-Abschätzvorgang tatsächlich Vokaltaktresonanzen aus der Erregungsinformation entfaltet, so daß das Restsignal relativ weniger der Vokaltraktresonanzen (Formanten) und relativ mehr der Erregungsinformation (Tonhöhe) beinhaltet. Jedoch weisen derartige, auf einem Rest basierende Tonhöhenabschätzungstechniken ihre eigenen Schwierigkeiten auf. Das LPC-Modell selbst wird normalerweise hochfrequentes Rauschen in das Restsignal einbringen, und Abschnitte von diesem hochfrequenten Rauschen können eine höhere spektrale Dichte aufweisen als die tatsächliche Tonhöhe, welche detektiert werden sollte. Eine Lösung für diese Schwierigkeit ist einfach, das Restsignal bei etwa 1000 Hz Tiefpaß zu filtern. Dies entfernt das hochfrequente Rauschen, entfernt jedoch auch die legitimierte Hochfrequenzenergie, welche in den stimmlosen Bereichen der Sprache vorhanden ist, und macht das Restsignal nahezu nutzlos für stimmhafte Entscheidungen.

Ein Hauptkriterium in Sprachnachrichtenanwendungen ist die Qualität der reproduzierten Sprache. Systeme nach dem Stand der Technik hatten in dieser Hinsicht viele Schwierigkeiten. Insbesondere beziehen sich viele dieser Schwierigkeiten auf Probleme eines genauen Detektierens der Tonhöhe und der Stimmhaftigkeit des eingegebenen bzw. Eingabesprachsignals.

Es ist typischerweise sehr leicht, eine Tonhöhenperiode auf dem Doppelten oder der Hälfte ihres Werts unkorrekt abzuschätzen. Beispielsweise garantiert, wenn Korrelationsverfahren verwendet werden, eine gute Korrelation bei einer Periode P eine gute Korrelation bei einer Periode $2P$, und bedeutet auch, daß es für das Signal wahrscheinlicher ist, eine gute Korrelation bei einer Periode $P/2$ zu zeigen.

Jedoch erzeugen solche Verdopplungs- und Halbierungsfehler eine sehr lästige Verminderung der Stimm- bzw. Sprachqualität. Beispielsweise wird ein fehlerhaftes Halbieren der Tonhöhenperiode dazu neigen, eine quietschende Stimme zu erzeugen, und ein fehlerhaftes Verdoppeln der Tonhöhenperiode wird dazu neigen, eine rauhe Stimme zu erzeugen. Darüber hinaus ist es

wahrscheinlich, daß ein Verdoppeln oder Halbieren einer Tonhöhenperiode intermittierend bzw. mit Unterbrechungen auftritt, so daß die synthetisierte Stimme dazu neigen wird, mit Unterbrechungen zu knacksen oder zu kratzen.

Bevorzugte Anordnungen verwenden ein adaptives Filter, um das Restsignal zu filtern. Durch Verwendung eines zeitveränderlichen Filters, welches einen einzelnen Pol beim ersten Reflexionskoeffizienten (k_1 der Spracheingabe) aufweist, wird das hochfrequente bzw. Hochfrequenz-Rauschen aus den stimmhaften Perioden der Sprache entfernt, jedoch wird die hochfrequente Information in den stimmlosen Sprachperioden zurückgehalten. Das adaptiv gefilterte Restsignal wird dann als die Eingabe für die Tonhöhenentscheidung verwendet.

Es ist notwendig, die hochfrequente bzw. Hochfrequenz-Information in den stimmlosen Sprachperioden zurückzuhalten bzw. beizubehalten, um bessere Stimmhaftigkeits/Stimmlosigkeits-Entscheidungen zu gestatten. D.h., die "stimmlose" Stimmhaftigkeitsentscheidung wird normalerweise vorgenommen, wenn keine starke Tonhöhe vorgefunden wird, d.h., wenn keine Korrelationsverzögerung des Restsignals einen hohen normalisierten Korrelationswert liefert. Jedoch kann, wenn nur ein tiefpaßgefilterter Abschnitt des Restsignals während stimmlosen Sprachperioden getestet wird, dieses teilweise bzw. Teilsegment des Restsignals unechte Korrelationen aufweisen. D.h., die Gefahr ist, daß das abgeschnittene Restsignal, welches durch das festgelegte Tiefpaßfilter nach dem Stand der Technik erzeugt ist, nicht genug Daten beinhaltet, um zuverlässig zu zeigen, daß keine Korrelation während stimmloser Perioden besteht, und die zusätzliche, durch die hochfrequente Energie der stimmlosen Perioden zur Verfügung gestellte Bandbreite notwendig ist, um zuverlässig die unechten Korrelationsverzögerungen auszuschließen, welche andernfalls gefunden werden könnten.

Eine Verbesserung in Tonhöhen- und Stimmhaftigkeitsentscheidungen ist besonders kritisch für Sprachnachrichtensysteme, ist jedoch auch für andere Anwendungen wünschenswert. Beispielsweise eine Worterkennungsvorrichtung, welche Tonhöheninformation mit einbezieht, würde natürlich ein gutes Tonhöhenabschätzverfahren erfordern. In ähnlicher Weise wird eine Tonhöheninformation manchmal zur Lautsprecher- bzw. Sprecherüberprüfung verwendet, insbesondere über eine Telefonleitung, wo eine hochfrequente Information teilweise verloren ist. Darüber hinaus wäre für zukünftige Weitbereichserkennungssysteme es wünschenswert, fähig zu sein, die syntaktische Information zu berücksichtigen, welche durch die Tonhöhe angegeben ist. In ähnlicher Weise wäre eine gute Analyse der Stimmhaftigkeit für einige fortschrittliche Spracherkennungssysteme, z.B. Sprache-zu-Text-Systeme wünschenswert.

Der erste Reflexionskoeffizient k_1 ist ungefähr auf das hoch/niederfrequente Energieverhältnis und ein Signal bezogen. Siehe R. J. McAulay, "Entwurf eines robusten Tonhöhenabschätzers maximaler Wahrscheinlichkeit für Sprache und zusätzliches Rauschen", Technische Notiz, 1979-28, Lincoln Labs, 11. Juni 1979. Für k_1 nahe zu -1 gibt es mehr niederfrequente Energie in dem Signal als hochfrequente Energie und umgekehrt für k_1 nahe zu 1 . Somit wird durch Verwendung von k_1 zum Bestimmen des Pols eines 1-poligen Deemphasis-Filters das Restsignal in den stimmhaften Sprachperioden tiefpaßgefiltert und wird in den stimmlosen Sprachperioden hochpaßgefiltert. Dies bedeutet, daß die Formanten-Frequenzen von einer Berechnung der Tonhöhe während der stimmhaften Perioden ausgeschlossen sind, während die notwendige Hochbandbreiteninformation in den stimmlosen Perioden zur genauen Detektion der Tatsache beibehalten wird, daß keine Tonhöhenkorrelation besteht.

* Vorzugsweise wird eine nachverarbeitende, dynamische Programmieretechnik verwendet, und nicht nur einen optimalen Tonhöhenwert, sondern auch eine optimale Stimmhaftigkeitsentscheidung zur Verfügung zu stellen. D.h., sowohl Tonhöhe wie auch Stimmhaftigkeit werden von Rahmen zu Rahmen nachgeführt und ein kumulativer Nachteil für eine Sequenz bzw. Abfolge von

akustische oder optische Signale, wie z. B. Radio, Fernseher oder Mobilfunkgerät. Erfindungsgemäß wird gebündelte modulierte elektromagnetische Strahlung in den Organismus des Empfängers derart gesendet, daß bei diesem Reaktionen ausgelöst werden, die einer beabsichtigten Gedankenübertragung entsprechen. Die Erfindung kann Anwendung finden z. B. (a) zur Unterstützung der Kommunikation mit Personen in Bunkern und verschütteten Personen nach Erdbeben, (b) zur unauffälligen Nachrichtenübertragung an Sicherheitspersonal, (c) zur Unterstützung von wichtigen Verhandlungen und in der Öffentlichkeit vortragenden Personen, (d) zur Sensibilisierung von Personen bezüglich wichtiger Themen in Notfallsituationen, z. B. durch Sendung von Gefahrenhinweisen, (e) in Kombination mit Detektionsmethoden zum Profiling und Gedankenlesen von Kriminellen, (f) zur Therapie und Prophylaxe von bestimmten pathologischen Beeinträchtigungen des Hirnstoffwechsels und zur Beeinflussung von bestimmten nichtpathologischen Limitierungen, Streßsituationen und Alterungsprozessen des Hirnstoffwechsels.

ANSPRÜCHE(33)

1. Richtfunkeinrichtung, dadurch gekennzeichnet, daß – die Richtfunkeinrichtung gebündelte modulierte elektromagnetische Strahlung erzeugt und zu einem menschlichen Empfänger sendet, – die Trägerfrequenz der gebündelten modulierten elektromagnetischen Strahlung zwischen 106 Hz (= 1 MHz) und 1014 Hz (= 100 THz) liegt, – eine Modulationsfrequenz der Trägerfrequenz zwischen 0,01 Hz und 1011 Hz (= 100 GHz) liegt, – die Entfernung zwischen der Richtfunkeinrichtung und dem Empfänger mehr als 10 m beträgt, – die gebündelte modulierte elektromagnetische Strahlung auf den Organismus des Empfängers auf solche

Weise einwirkt, daß mit erheblicher Wahrscheinlichkeit eine beabsichtigte Änderung der Gedanken oder Handlungen des Empfängers erzeugt wird, – die Änderung der Gedanken oder Handlungen des Empfängers mit wissenschaftlichen Methoden nachweisbar ist, – die Sendung der gebündelten modulierten elektromagnetischen Strahlung vom Empfänger selbst nicht bewußt wahrgenommen wird, – der Informationsgehalt der Sendung der gebündelten modulierten elektromagnetischen Strahlung mehr als 100 bit umfaßt, – der Empfänger zum Empfang von mittels der gebündelten modulierten elektromagnetischen Strahlung gesendeten Information keine elektronischen Hilfsmittel benötigt, die eine Umwandlung der elektromagnetischen Strahlung in akustische oder optische oder mechanische Signale oder Geruchssignale oder Geschmackssignale bewirken.

2. Richtfunkeinrichtung nach Anspruch 1, dadurch gekennzeichnet, daß zusätzlich auch vom Empfänger bewußt wahrnehmbare Signale gesendet werden.

3. Richtfunkeinrichtung nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die Sendung der gebündelten modulierten elektromagnetischen Strahlung beim Empfänger mindestens einen der folgenden fünf auf der Wirkung der elektromagnetischen Strahlung beruhenden Effekte verursacht: (i) unterschwellige Signale im Bereich von 12 Hz – 25 kHz, (ii) wahrnehmbare Signale im Bereich von 12 Hz – 25 kHz, (iii) unterschwellige Signale mit Frequenzen unterhalb 12 Hz, (iv) unterschwellige Signale mit Frequenzen oberhalb 25 kHz, (v) wahrnehmbare Signale mit Frequenzen außerhalb des Bereichs 12 Hz – 25 kHz.

4. Richtfunkeinrichtung nach Anspruch 1, dadurch gekennzeichnet, daß Laute einer Sprache in eine Sequenz von Impulsen umgewandelt und diese Sequenz der elektromagnetischen Strahlung aufmoduliert wird.

5. Richtfunkeinrichtung nach Anspruch 1, dadurch gekennzeichnet, daß sie eine Kamera oder andere Detektionseinrichtung involviert, die bei der Trägerfrequenz der gebündelten modulierten elektromagnetischen Strahlung zur Gedankenübertragung sensitiv ist.
6. Richtfunkeinrichtung nach Anspruch 1, dadurch gekennzeichnet, daß sie einen Computer involviert, der zu sendende Stimuli für eine beabsichtigte Gedankenübertragung berechnet.
7. Richtfunkeinrichtung nach Anspruch 1, dadurch gekennzeichnet, daß die gebündelte modulierte elektromagnetische Strahlung zu mehr als 50% aus einer Quelle mit induzierter Emission von Strahlung stammt.
8. Richtfunkeinrichtung nach Anspruch 1, dadurch gekennzeichnet, daß die gebündelte modulierte elektromagnetische Strahlung zu mehr als 50% aus einem Maser, Laser, Phased Array, Diodenbündel, Magnetron oder Klystron stammt.
9. Richtfunkeinrichtung nach Anspruch 1, dadurch gekennzeichnet, daß die gebündelte modulierte elektromagnetische Strahlung zu mehr als 50% in weniger als $1 \text{ Grad} \times 1 \text{ Grad}$ Raumwinkel abgestrahlt wird.
10. Richtfunkeinrichtung nach Anspruch 1, dadurch gekennzeichnet, daß die Entfernung zwischen der Richtfunkeinrichtung und dem Empfänger mehr als 1 km beträgt.
11. Richtfunkeinrichtung nach Anspruch 1, dadurch gekennzeichnet, daß die gebündelte modulierte elektromagnetische Strahlung auf den Organismus des Empfängers auf solche Weise einwirkt, daß mit mehr als 5% Wahrscheinlichkeit eine beabsichtigte Änderung der Gedanken oder Handlungen des Empfängers erzeugt wird.
12. Richtfunkeinrichtung nach Anspruch 1, dadurch gekennzeichnet, daß die gebündelte modulierte elektromagnetische Strahlung auf den Organismus des Empfängers auf solche Weise einwirkt, daß mit mehr als 95% Wahrscheinlichkeit eine beabsichtigte Änderung der Gedanken oder Handlungen des Empfängers erzeugt wird.
13. Richtfunkverfahren, dadurch gekennzeichnet, daß – gebündelte modulierte elektromagnetische Strahlung erzeugt und zu einem menschlichen Empfänger gesendet wird, – die Trägerfrequenz der gebündelten modulierten elektromagnetischen Strahlung zwischen 106 Hz (= 1 MHz) und 1014 Hz (= 100 THz) liegt, – eine Modulationsfrequenz der Trägerfrequenz zwischen 0,01 Hz und 1011 Hz (= 100 GHz) liegt, – die Entfernung zwischen der Richtfunkeinrichtung und dem Empfänger mehr als 10 m beträgt, – die gebündelte modulierte elektromagnetische Strahlung auf den Organismus des Empfängers auf solche Weise einwirkt, daß mit erheblicher Wahrscheinlichkeit eine beabsichtigte Änderung der Gedanken oder Handlungen des Empfängers erzeugt wird, – die Änderung der Gedanken oder Handlungen des Empfängers mit wissenschaftlichen Methoden nachweisbar ist, – die Sendung der gebündelten modulierten elektromagnetischen Strahlung vom Empfänger selbst nicht bewußt wahrgenommen wird, – der Informationsgehalt der Sendung der gebündelten modulierten elektromagnetischen Strahlung mehr als 100 bit umfaßt, – der Empfänger zum Empfang von mittels der gebündelten modulierten elektromagnetischen Strahlung gesendeten Information keine elektronischen Hilfsmittel benötigt, die eine Umwandlung der elektromagnetischen Strahlung in akustische oder optische oder mechanische Signale oder Geruchssignale oder Geschmackssignale bewirken.
14. Richtfunkverfahren nach Anspruch 13 dadurch gekennzeichnet, daß zusätzlich auch vom

Empfänger bewußt wahrnehmbare Signale gesendet werden.

15. Richtfunkverfahren nach Anspruch 13 oder 14 dadurch gekennzeichnet, daß die Sendung der gebündelten modulierten elektromagnetischen Strahlung beim Empfänger mindestens einen der folgenden fünf auf der Wirkung der elektromagnetischen Strahlung beruhenden Effekte verursacht: (i) unterschwellige Signale im Bereich von 12 Hz – 25 kHz, (ii) wahrnehmbare Signale im Bereich von 12 Hz – 25 kHz, (iii) unterschwellige Signale mit Frequenzen unterhalb 12 Hz, (iv) unterschwellige Signale mit Frequenzen oberhalb 25 kHz, (v) wahrnehmbare Signale mit Frequenzen außerhalb des Bereichs 12 Hz – 25 kHz.

16. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß Laute einer Sprache in eine Sequenz von Impulsen umgewandelt und diese Sequenz der elektromagnetischen Strahlung aufmoduliert wird.

17. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß sie eine Kamera oder andere Detektionseinrichtung involviert, die bei der Trägerfrequenz der gebündelten modulierten elektromagnetischen Strahlung zur Gedankenübertragung sensitiv ist.

18. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß sie einen Computer involviert, der zu sendende Stimuli für eine beabsichtigte Gedankenübertragung berechnet.

19. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß die gebündelte modulierte elektromagnetische Strahlung zu mehr als 50% aus einer Quelle mit induzierter Emission von Strahlung stammt.

20. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß die gebündelte modulierte elektromagnetische Strahlung zu mehr als 50% aus einem Maser, Laser, Phased Array, Diodenbündel, Magnetron oder Klystron stammt.

21. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß die gebündelte modulierte elektromagnetische Strahlung zu mehr als 50% in weniger als $1 \text{ Grad} \times 1 \text{ Grad}$ Raumwinkel abgestrahlt wird.

22. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß die Entfernung zwischen der Richtfunkeinrichtung und dem Empfänger mehr als 1 km beträgt.

23. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß die gebündelte modulierte elektromagnetische Strahlung auf den Organismus des Empfängers auf solche Weise einwirkt, daß mit mehr als 5% Wahrscheinlichkeit eine beabsichtigte Änderung der Gedanken oder Handlungen des Empfängers erzeugt wird.

24. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß die gebündelte modulierte elektromagnetische Strahlung auf den Organismus des Empfängers auf solche Weise einwirkt, daß mit mehr als 95% Wahrscheinlichkeit eine beabsichtigte Änderung der Gedanken oder Handlungen des Empfängers erzeugt wird.

25. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß eine auf der Wirkung von modulierter Mikrowellenenergie basierende Gefühlsbeeinflussung involviert ist.

26. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß Gedankenübertragung an eine Zielperson durch Gegenstände aus Beton, Stein, Plastik oder Holz hindurch erfolgt.

27. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß Gedankenübertragung an eine Zielperson über mehr als 10 km Entfernung erfolgt.

28. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß der Empfänger eine Einrichtung zur Verstärkung des Gedankensignals benutzt, z.B. eine Antenne oder einen Mikrowellenverstärker.

29. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß der Empfänger vom Sender mittels einer Kamera beobachtet wird und die Trägerfrequenz der gebündelte modulierte elektromagnetische Strahlung zur Gedankenübertragung eine Frequenz ist, bei der die Kamera zur Beobachtung des Empfängers sensitiv ist.

30. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß die zur beabsichtigten Gedankenübertragung notwendigen Signale computerbasiert unter

Ausnutzung eines Satzes von Korrelationen zwischen Stimuli und Reaktionen vorhergesagt werden.

31. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß die Gedankenübertragung direkt von Sender zu Empfänger oder über eine Bündlungseinrichtung oder Verstärkereinrichtung oder Relaisstation erfolgt.

32. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß der Trägerfrequenz eine Zwischenfrequenz aufmoduliert ist, der das Nutzsignal aufmoduliert ist.

33. Richtfunkverfahren nach Anspruch 13, dadurch gekennzeichnet, daß Worte in Impulsfolgen, deren Einhüllende dem Intensitätsverlauf der Worte entspricht, umgeformt und computergespeichert werden und vom Computer abgerufene Impulsfolgen dem elektromagnetischen Strahl aufmoduliert und mit so geringen Intensitäten gesendet werden, daß der Empfänger die Übertragung nicht bewußt wahrnimmt.

BESCHREIBUNG

* Hintergrund der Erfindung

* Anwendungsgebiet der Erfindung

* [0001]

Die Erfindung betrifft langreichweitige Gedankenübertragung und langreichweitiges Gedankenlesen. Anwendungen sind z.B. die Erweiterung der herkömmlichen Kommunikationsmittel, die Unterstützung öffentlicher Auftritte wichtiger Persönlichkeiten und wichtiger Verhandlungen, die Sendung von wichtigen Gefahrenhinweisen in Notsituationen, die aktive Abwendung von erheblichen Gefahren, die Untersuchung von Kriminellen, die Unterstützung der Hirnforschung. Dabei werden Limitierungen herkömmlicher Methoden der Informationsübertragung, wie z.B. Mobilfunktelefon, Radio und Fernsehen, überwunden.

* Charakteristik des bekannten Standes der Technik

* [0002]

Bei modernen Medien, wie z.B. Radio und Fernsehen, wird ein elektronisches Gerät benötigt, das elektromagnetische Strahlung in ein wahrnehmbares akustisches oder optisches Signal umwandelt

und einzelne Personen sind i.a. nicht individuell mit Informationen zu versorgen. Bekannt ist auch z.B. (a) der Effekt der Hörbarkeit von bestimmten RADAR-Impulsen (Beobachtungen im II. Weltkrieg), (b) die direkte akustische Wahrnehmbarkeit ("Hörbarkeit") von modulierter Mikrowellenenergie bei Einstrahlung in den Kopf (1; Frey, 1961; Frey, 1962; Frey & Messener, 1973; Lin, 1978; Frey & Corin, 1979; Brunkan, 1989; Lin, 1989; Stocklin, 1989; Frey, 1993), (c) die Gefühlssteuerung mittels akustischer oder elektrischer Stimulierung (Meland, 1980; Gall, 1994), und (d) die Anwendung akustischer Signale zur unterschwelligem Beeinflussung (Lowery, 1992). Die akustischen Wahrnehmungen bei Einwirkung von gepulsten Mikrowellenstrahlen beruhen unter den meisten bisher gewählten experimentellen Bedingungen auf der Erzeugung thermoelastischer Druckwellen im Innenohr (Lin, 1989).

* [0003]

Der menschliche Körperdipol hat bei 1,80 m Körperlänge eine Resonanzfrequenz von 80 MHz. Die individuell etwas verschiedenen elektromagnetischen Resonanzfrequenzen des menschlichen Kopfes liegen um 400 MHz bei Erwachsenen und um 700 MHz bei Kleinkindern (Lin, 1989). Aufgrund des Skineffekts ist die Eindringtiefe von elektromagnetischer Strahlung in den Organismus frequenzabhängig, z.B. bei einer Einstrahlung auf den Kopf erfolgt die Absorption bei 2,5 GHz Frequenz hauptsächlich in den äußeren 1–2 cm des Hirns, dagegen bei 900 MHz mehr im Inneren des Hirns (Lin, 1989).

* [0004]

Bekannt sind auch elektromagnetische Waffen, mit denen (bei Beobachtung mittels Millimeterwellenteleskopen oder Mikrowellendetektoren) über größere Entfernungen oder durch nichtmetallische Wände hindurch Menschen betäubt oder ausgeschaltet werden können.

* [0005]

Bekannt ist auch die unterschwellige Stimulierung mit herkömmlichen akustischen Verfahren. Beispielsweise können aufmodulierte Rhythmen bei 1,7 – 3,5 Hz zur Förderung von Schlafbedürfnis dienen. Abnormale Zustände des Bewußtseins lassen sich durch Rhythmen im Bereich von 3,5 – 7 Hz und 28 – 56 Hz fördern. Der normale Rhythmus des menschlichen Hirns liegt bei 7 – 14 Hz und bei 14 – 28 Hz im Falle der Erregung oder Angst (Gall, 1994).

* [0006]

Die Ideen von Gedankenübertragung und Gedankenlesen werden jedoch gewöhnlich als nichtpraktikable Phantasien angesehen (siehe z.B. Chapman, 1998) und mit keinem der genannten Systeme allein können langreichweitige Gedankenübertragungen oder gar Gedankenlesen effizient realisiert werden, z.B. über eine Entfernung von einigen Kilometern. Menschen, die behaupten, ohne technische Hilfsmittel Gedanken über große Entfernungen senden oder empfangen zu können (z.B. einige Esoteriker), konnten bisher keinen Wirkungsnachweis führen. Auch belegen zahlreiche utopische Filme mit Episoden von fernreichender Gedankenübertragung oder Gedankenlesen, daß es bisher für diesen Wunschtraum keine praktikable Lösung mit guter Effizienz gibt.

* Ziel der Erfindung

* [0007]

Das Ziel der Erfindung ist die Erweiterung der Möglichkeiten moderner Medien in Form der langreichweitigen Gedankenübertragung die seitens des Empfängers keine elektronischen

Hilfsmittel wie z.B. Radio, Fernseher oder Mobilfunktelefon, benötigt.

* Literatur

- o Brunkan, W.B. (1989) Hearing system. US-Patent 4.877.027.
- o Chapman, R.K. (1998) Mental telepathy debunked: counter-arguments against the concept of thought transmission und mind-reading ideas. ISBN = 0-9698637-6-4.
- o Frey, A.H. (1961) Auditory system response to modulated electromagnetic energy. Aerospace Med. 32, 1140-1142.
- o Frey, A.H. (1962) Human auditory system response to modulated electromagnetic energy. J. Appl. Physiol. 17, 689-692.
- o Frey, A.H & Messener, R. (1973) Human perception of illumination with pulsed UHF electromagnetic energy, Science 181, 356-358.
- o Frey, A.H & Corin, E. (1979) Holographic assessment of a hypothesized microwave hearing mechanism. Science 206, 232-234.
- o Frey, A.H. (1993) Electromagnetic field interactions with biological systems. FASEB Journal 7, 272-281.
- o Gall, J. (1994) Method and system for altering consciousness. US-Patent 5.289.438; und Referenzen darin.
- o Lin, J.C. (1978) Microwave auditory effects and applications. Charles C. Thomas, Publisher, Springfield, IL, USA.
- o Lin, J.C. (1989) Electromagnetic interaction with biological systems. Plenum Press, New York.
- o Lowery, O.M. (1992) Silent subliminal presentation system. US-Patent 5.159.703.
- o Meland, B.C. (1980) Apparatus for electrophysiological stimulation. US-Patent 4.227.516.,
- o Stocklin, P.L. (1989) Hearing device. US-Patent 4.858.612.

* Detaillierte Beschreibung der Erfindung

* [0008]

Der Erfindung liegt die Aufgabe zugrunde, bestimmte erwünschte Nachrichtenübertragungen zu ermöglichen ohne an die Einschränkungen von herkömmlich verwendeten elektronischen Mitteln gebunden zu sein. Erfindungsgemäß wird die Aufgabe durch die Verwendung von fernreichender Gedankenübertragung gelöst, wobei die Gedankenübertragung auf Richtfunk beruht. Im Gegensatz zum herkömmlichen Richtfunk wird jedoch der elektromagnetische Strahl (Gedankenstrahl) direkt in den Organismus des Empfängers eingekoppelt, z.B. in den Kopf, die Großhirnrinde, das Innenohr, die Gehörnerven oder Sehnerven. In Abhängigkeit von speziellen in den elektromagnetischen Strahl eingebrachten Signalen (z.B. mittels Amplitudenmodulation) bewirkt

diese Einkopplung beim Empfänger eine beabsichtigte Änderung der Gedanken. Im allgemeinen ist die Änderung der Gedanken des Empfängers nur statistisch wirksam, d.h. es wird lediglich die Wahrscheinlichkeit für bestimmte Gedanken auf beabsichtigte Weise erhöht oder verringert. Im Einzelfall kann die Änderung jedoch auch determiniert sein. Die Gedankenübertragung eignet sich in einigen Anwendungsfällen zur Kombination mit Beobachtungen mittels Millimeterwellenkameras und mikrowellen-basierten Sprachübertragungen, in der die Hörbarkeit von modulierter Mikrowellenenergie genutzt wird, kann jedoch auch unabhängig davon betrieben werden.

* [0009]

Z.B. in einer einfachen Ausführung eines Gedankenübertragungsgerätes spricht der Betreiber des Geräts (Beobachter, Observer) den zu sendenden Gedanken in ein Mikrofon, das elektrische Signal des Mikrophons wird mittels einer Elektronik in eine Folge von Impulsen umgewandelt (z.B. Rechteckimpulse von 100 Mikrosekunden Dauer mit 200 Mikrosekunden Abstand; gegebenenfalls werden Impulsfolgen computergespeichert und von dort nach Bedarf abgerufen), die Folge von Impulsen wird dein Mikrowellenstrahl aufmoduliert, der an den Empfänger gesendet wird und solch eine geringe Intensität hat, daß der Empfänger keine bewußte Wahrnehmung der Sendung hat, sondern diese nur unterschwellig wirkt. Anstelle der Impulsfolge kann auch ein flankenversteilertes Signal (z.B. mittels mehrfacher Quadrierung) oder das ursprüngliche Signal genutzt werden.

* [0010]

Z.B. in einer komplizierteren Ausführung eines Gedankenübertragungsgerätes gibt der Betreiber (Beobachter, Observer) des Geräts den zu sendenden Gedanken in einen Computer ein (oder einen anderen Überträger), der mit Hilfe von Tabellen oder neuronalen Netzen den zu sendenden Gedanken in eine Sequenz von Signalen übersetzt, die dem Mikrowellenstrahl, der an den Empfänger gesendet wird, aufmoduliert wird. Diese Sequenz von Signalen kann mikrowelleninduzierte bewußt wahrnehmbare akustische Signale (z.B. Klickgeräusche, Rhythmen, Sprache, Musik) und mikrowelleninduzierte nur unbewußt wahrnehmbare akustische Signale (z.B. Klickgeräusche, Rhythmen, Sprache, Musik) und mikrowelleninduzierte niederfrequente elektrisch wirksame Rhythmen enthalten. Die Berechnung der Übersetzungstabellen zwischen zu sendenden Gedanken und Sequenz von Signalen erfolgt z.B. unter Ausnutzung eines Satzes von Korrelationen zwischen Stimuli und Reaktionen. Das Training der neuronalen Netze zur Übersetzung zwischen zu sendenden Gedanken und Sequenz von Signalen erfolgt z.B. unter Beobachtung der Reaktionen auf einen Satz von Stimuli.

* Frequenzen

* [0011]

Zur besseren Wirksamkeit der Gedankenübertragung können als Trägerfrequenz des elektromagnetischen Strahls oder als auf die Trägerfrequenz aufmodulierte Frequenzen bestimmte Resonanzen von Körperteilen (z.B. des Kopfes, Teilen des Innenohrs, Sehnerv) gewählt werden. Z.B. eignen sich Trägerfrequenzen und eventuell der Trägerfrequenz aufmodulierte Zwischenfrequenzen um 80 MHz bzw. 400 – 700 MHz bzw. 1 – 100 GHz zur Adressierung des Körpers bzw. des Kopfes bzw. von Organteilen (z.B. Innenohr, Nerven).

* [0012]

Für die aufmodulierten Signale eignet sich vor allem der Frequenzbereich von 1 Hz – 1 GHz. Beispielsweise können die Frequenzen bei Sprachsignalen (bewußt wahrnehmbare oder

unterschwellig wirksame) im Bereich von 16 Hz – 20 kHz liegen, aber z.B. bei Transformation

in eine Impulsfolge weit darüber, z.B. im MHz-Bereich. Besonders niedrige Frequenzen eignen sich beispielsweise zur Beeinflussung des Bewußtseinszustandes und zur Gefühlsbeeinflussung. Beispielsweise können in Analogie zur herkömmlichen akustischen Stimulierung aufmodulierte Rhythmen bei 1,7 – 3,5 Hz bzw. 3,5 – 7 Hz und 28 – 56 Hz zur Förderung von Schlafbedürfnis bzw. veränderter Bewußtseinszustände dienen.

* Modulation

* [0013]

Für die Modulation des elektromagnetischen Strahls gibt es verschiedene Möglichkeiten, die einzeln oder kombiniert angewendet werden können, z.B. (a) Laute einer Sprache oder andere Signale werden in eine Impulsfolge umgewandelt, die den elektromagnetischen Strahl aufmoduliert wird, oder (b) Laute einer Sprache oder andere Signale werden direkt auf den elektromagnetischen Strahl aufmoduliert. Die Sendung erfolgt wahrnehmbar oder nicht wahrnehmbar – abhängig z.B. von Intensität, Art der Modulation, Ort der Einstrahlung in den Organismus und Frequenz.

* Strahlungsquellen

* [0014]

Für die Generierung des elektromagnetischen Strahls (Gedankenstrahl) eignen sich vor allem MASER (Microwave Amplification by Stimulated Emission of Radiation) und LASER (Light Amplification by Stimulated Emission of Radiation), wobei die Wellenlänge nicht notwendigerweise im klassischen Mikrowellenbereich (300 MHz – 300 GHz) liegen muß (2). Erfindungsgemäß sind auch Strahlungsquellen, die induzierte Emission von elektromagnetischer Strahlung involvieren, die außerhalb des klassischen Mikrowellenbereichs liegt, mit einbezogen. Insbesondere stehen MASER an allen Stellen der "Detaillierten Beschreibung der Erfindung", den Ausführungsbeispielen und Abbildungen samt Legenden synonym für Strahlungsquellen mit induzierter Emission, wie z.B. MASER und LASER (z.B. Freie-Elektronen-LASER). Weiterhin kommen als Strahlungsquellen auch Magnetrons, Gyrotrons, Klystrons, Halbleiterdioden und Phased Arrays in Betracht.

* Sendeleistung

* [0015]

Je nach Größe der Übertragungsverluste kann die Sendeleistung pro adressierter Person im Bereich der Leistung des menschlichen Hirns (ca. 40 W) oder etwas höher liegen, aber bei z.B. Einstrahlung in das Innenohr oder gar die Nervenenden der Sinnesorgane auch je nach Anwendung deutlich niedriger sein. Z.B. bei Transmissionen über mehrere Kilometer Entfernung durch Gebäudewände hindurch können auch Sendeleistungen von über 1000 W je adressierter Person erforderlich sein, um die Transmissionsverluste auszugleichen. Spezielle Maßnahmen können erforderlich sein, um Waffenwirkung auf Personen im Strahlengang zu vermeiden (eine energiereiche Einstrahlung kann betäubend und kurzzeitige Temperaturerhöhung des Hirns über 45°C kann tödlich sein). Andererseits kann bei Abwesenheit von wesentlichen Übertragungsverlusten eine Leistung von wesentlich weniger

als durchschnittlich 1 W für eine unterschwellige Gedankenübertragung ausreichen. Da excessive Absorption von Mikrowellen im Gewebe Schäden verursachen kann (v.a. bei sich schnell teilenden

Zellen und Neuronen), wird man in vielen Anwendungen geringe Strahlungsgesamtenergien bevorzugen.

* Automatisierung

* [0016]

Die Gedankenübertragung kann z.B. vom Gedankenübertragungsgerät zu Mensch vollautomatisch oder teilautomatisch oder von Mensch zu Mensch mit zwischengeschaltetem Gedankenübertragungsgerät erfolgen.

* 1. Ausführungsbeispiel

* [0017]

Auf ein Fahrzeug montiertes Gedankenübertragungsgerät, welches einen fokussierten Mikrowellenstrahl erzeugt, auf geeignete Weise moduliert und zum Empfänger (Target) sendet (3). Das Gesamtgewicht des Gedankenübertragungsgeräts mit MASER (Maser) zur Erzeugung des elektromagnetischen Strahls (Beam), Mikrofon zur Eingabe von Sprachsignalen durch den Beobachter (Headset), aufladbarer Energiequelle zur Pufferung von Stromfluktuationen und Detektor (Detector) zur Beobachtung und Unterstützung der Strahlnachführung kann z.B. 100 kg betragen. Um für den elektromagnetischen Strahl eine gute Bündelbarkeit bei noch ausreichender Durchdringung von Luft, Wänden und Erde zu ermöglichen, eignet sich als Trägerfrequenz des MASERs z.B. der Bereich von 1 – 1000 GHz. Die Trägerfrequenz des elektromagnetischen Strahls (Beam) zur Gedankenübertragung kann z.B. eine Frequenz sein, bei der der Detektor (Detector) zur Beobachtung des Empfängers sensitiv ist. Gedankenübertragung und Beobachtung des Empfängers erfolgen z.B. über größere Entfernungen durch Luft oder durch Wände aus Beton, Stein, Plastik oder Holz.

* [0018]

Die Gedankenübertragung erfolgt z.B. indem der Beobachter den Strahl des MASERs (Beam) auf den Kopf des Empfängers (Target) richtet und in das Mikrofon spricht, wobei das elektrische Signal des Mikrophons mittels der Elektronik des Gedankenübertragungsgeräts der Trägerfrequenz des MASERs auf geeignete Weise (z.B. in Form einer Impulsfolge, deren Amplitude mit der Amplitude des elektrischen Signals des Mikrophons korreliert) aufmoduliert wird und wobei die MASER-Strahlung Spannungen im Kopf des Empfängers induziert, was beim Empfänger z.B. als unterschwelliges Signal wirkt. Alternativ kann die Modulation des Mikrophonsignals auf die Trägerfrequenz z.B. unter Verwendung einer elektronischen Übersetzungseinrichtung erfolgen, die z.B. zuvor unter Ausnutzung eines Satzes von Korrelationen zwischen Stimuli und Reaktionen trainiert wurde. Alternativ kann die zu sendende Gedankensequenz in einen Computer eingegeben werden, der das zu sendende Signal berechnet. Zur Entwicklung des Programms zur Übersetzung der zu sendenden Gedanken in die auf den elektromagnetischen Strahl (Beam) aufzumodulierenden Sequenzen können z.B. eine große Zahl von schwachen Korrelationen zwischen Gedanken und Stimuli

genutzt worden sein. Das Computerprogramm kann z.B. ein neuronales Netz (4) enthalten, welches zuvor mit z.B. einem großen Satz von Paaren von Stimuli und Reaktion trainiert wurde und nach dem Training beabsichtigte Gedanken auf Sätze von Stimuli projiziert. Die aufmodulierten Sequenzen müssen nicht im hörbaren Frequenzbereich liegen. Beispielsweise können u.a. auch niederfrequente Signale im Bereich von 1–20 Hz auf die Trägerfrequenz des MASERs aufmoduliert werden, was zu einer Beeinflussung des Empfängers führen kann. Auch Signale im Bereich über 20

kHz sind verwendbar. In vielen Anwendungsfällen – insbesondere wenn man in die eigenständigen Handlungen der Empfänger nicht zu sehr eingreifen möchte – wird man sich mit einer unauffälligen und für die Empfänger unbewußten Veränderung der Wahrscheinlichkeiten gestimmter Gedanken begnügen.

* 2. Ausführungsbeispiel

* [0019]

Handgetragenes Gedankenübertragungsgerät, das einem MASER (Maser), ein Mikrophon (Headset) zur Eingabe der Sprachsignale durch den Beobachter (Observer), eine aufladbare Energiequelle (Battery) und zur Beobachtung einen Detektor (Detector), z.B. eine Millimeterwellenkamera, enthält (5). Das Gedankenübertragungsgerät kann an das Stromfestnetz, das Stromnetz eines Fahrzeugs oder einen Generator (Power generator) mit z.B. 200 W Leistung angeschlossen werden. Mittels der Anzeige (Display) und des Handgriffs (Handle) wird das Gedankenübertragungsgerät zum Empfänger (Target) nachgeführt. Verschiedene Schalter (switches) und die Elektronik (Electronics) erlauben die Einstellung verschiedener Modi wie z.B. Sendung eingespeicherter Signale, automatische Intensitätsanpassung, Art der Modulation zur Übertragung der Sprachsignale des Beobachters. Das Gedankenübertragungsgerät kann mittels eines Verbindungselements (Connector) auf Stative oder Fahrzeuge beweglich montiert werden. Die Gedankenübertragung erfolgt z.B. durch die Sendung von vorher ermittelten Sequenzen. Im Gegensatz zu betäubenden Schüssen mit elektromagnetischen Waffen wird mit vergleichsweise geringen Intensitäten gearbeitet. Unterhalb der Intensität bewußter Wahrnehmung wirkt die elektromagnetische Strahlung, auf die z.B. ein akustisches Signal aufmoduliert ist, unbewußt als unterschwelliges scheinbar akustisches Signal und beeinflusst die Gedanken des Empfängers. Bei höheren Intensitäten ist der elektromagnetische Strahl direkt fühlbar. Neben hörbarer und unterschwelliger Sprache, Musik und Rhythmen lassen sich z.B. auch niederfrequente Rhythmen (z.B. unter 16 Hz) und Signale im Bereich oberhalb von 20 kHz auf den elektromagnetischen Strahl aufmodulieren.

* 3. Ausführungsbeispiel

* [0020]

Auf ein Fahrzeug (6), einen Sendeturm (7), ein Haus (8) oder in ein Flugobjekt (9) montiertes (gegebenenfalls bewegungsstabilisiertes) Gedankenübertragungsgerät mit einer Quelle intensiver elektromagnetischer Strahlung und einer Einrichtung zur Modulation der Strahlung entsprechend den zu sendenden Gedanken, z.B. einem Computer, der für die zu sendenden Gedanken eine Sequenz von elektromagnetischen Reizen berechnet (z.B. unterschwellig oder bewußt wahrnehmbare Sprache, Musik, Rhythmen und Lautfolgen, die gleichzeitig oder

sequentiell gesendet werden). Zur Gedankenübertragung mittels erdnaheer Satelliten (10) weist der MASER (Maser) ein besonders kleinen Öffnungswinkel auf. Gedanken werden z.B. erzeugt durch Ausnutzen von vielen schwachen Korrelationen zwischen Gedanken und Sätzen von Stimuli. Bei einer Anwendung über einen langen Zeitraum können eine große Zahl von Korrelationen gemessen werden und kann die Ausnutzung von relativ schwache Korrelationen zwischen Stimuli und Gedanken zu einer erheblichen Veränderung der Wahrscheinlichkeit bestimmter Gedanken führen. Um für stark gebündelten Richtfunk geeignete hohe Trägerfrequenzen zur Übertragung niederfrequenter Gedankensignale zu nutzen, wird das zu sendende Signal auf die Trägerfrequenz des Richtfunkstrahls aufmoduliert, z.B. mittels Amplitudenmodulation. Wenn das aufmodulierte Signal ein akustisches Signal ist (z.B. eine Amplitudenmodulation mit einer hörbaren Frequenz vorliegt), kann oberhalb einer bestimmten Intensität diese modulierte elektromagnetische Strahlung

direkt als scheinbar akustisches Signal gehört werden. Zur Verringerung der zur Gedankenübertragung notwendigen Intensitäten der elektromagnetischen Strahlung lassen sich Einstrahlungen in einzelne Nervenbündel nutzen, z.B. Hörnerven und Sehnerven. Das kann nicht nur über die Nutzung von deren Resonanzfrequenzen erfolgen, sondern auch durch Einstrahlung mit so hoher Präzision, daß diese Organteile vom Strahl bevorzugt getroffen werden.

* 4. Ausführungsbeispiel

* [0021]

Gedankenübertragung an Empfänger im Katastrophenfall (11). Gedankenübertragung kann in wichtigen Ausnahmesituationen zur Schadensbegrenzung und schnellen unkomplizierten Steuerung von Rettungsmaßnahmen hilfreich sein. Teilkomponenten der Gedankenübertragung können mikrowellen-gestützte Sprachübertragungen und Gefühlsbeeinflussungen der Empfänger sein. Stimuli können z.B. Sprache, Musik, Rhythmen und Lautfolgen sein. Die Stimulierung kann unterschwellig (d.h. unbewußt wahrnehmbar) oder bewußt wahrnehmbar sein. Mehrere Stimuli können gleichzeitig oder sequentiell gesendet werden, um eine bestimmte Reaktion auszulösen. Beispielsweise wird die Sendung bewußt wahrnehmbarer Wortteile mit der Sendung unterschwellig wirkender Rhythmen kombiniert. Die Gedankenübertragung hat z.B. eine beabsichtigte Änderung der Gedankenwelt des Empfängers, z.B. Motivierung zu schadensbegrenzenden Handlungen, zur Folge.

* 5. Ausführungsbeispiel

* [0022]

Mensch-zu-Mensch Gedankenübertragung: Das zu sendende Signal wird direkt vom Kopf einer Person abgegriffen und direkt oder in verarbeiteter Form (z.B. mittels Frequenzanalyse und Selektion der vorherrschenden Frequenz) auf den elektromagnetischen Strahl aufmoduliert. Aus diese Weise werden, z.B. Spannungs- oder Entspannungszustände, die sich durch unterschiedliche Frequenzen der Hirnaktivität unterscheiden, übertragen. Sender oder Empfänger können z.B. Personen im Wachkoma oder blinde Taubstumme sein.

* 6. Ausführungsbeispiel

* [0023]

Profiling und Gedankenlesen bei einem verurteilten Kriminellen im Rahmen des gesetzlich und sittlich zulässigen. Ein einfaches Verfahren wäre, der Person überraschend unterschwellig ein Schlüsselwort zu senden, welches nur für sie wichtige Bedeutung hat und mittels gleichzeitiger Beobachtung der Reaktion wird ein Verdacht erhärtet oder erweicht. Der Sendung des Schlüsselwortes kann eine Vorbereitungsphase (Sensibilisierungsphase) vorausgehen, in der z.B. die Gedanken der Person durch unterschwellige Signale auf das Schlüsselereignis gerichtet werden. Die computergestützte Gedankenübertragung ermöglicht jedoch weitaus höherentwickelte Methoden: Z.B. können bestimmte Schlüsselinformationen unterschwellig über einen längeren Zeitraum mit wechselnden Intensitäten gesendet und die Reaktionen des Empfängers mit dem Signal korreliert werden.

* 7. Ausführungsbeispiel

* [0024]

Verurteilte Kriminelle zur Abwehr von Gefahren unauffällig manipulieren oder ausforschen – soweit gesetzlich und sittlich zulässig (Abb. 12). Das zeitweilige Ausschalten aller Kriminellen mittels amplitudenmodulierter intensiver Mikrowellenstrahlen bei der Erstürmung eines Objekts (unauffällig durch Wände hindurch) hat gewisse Risiken des Fehlschlags und ist bei elektromagnetisch abgeschirmten Objekten schwierig. Die Gedankenübertragung ermöglicht, diese Risiken zu verringern. In lebensbedrohlichen Situationen kann es akzeptabel sein, eine Gedankenmanipulation auf Nicht-Kriminelle beteiligte Personen auszuweiten, was die Anwendung auf abgeschirmte Objekte vereinfacht (z.B. Strahlung durch Löcher in der Abschirmung diffus in den gesamten Innenraum). Beispielsweise kann die Strahlungsleistung für eine Gedankenübertragung unter 1/1000 der für das zeitweilige Betäuben der Kriminellen notwendigen Strahlungsleistung liegen, was auch ein erheblicher Kostenfaktor sein sollte. Ein weiterer Vorteil ist, daß die Hardware der Gedankenübertragung leicht auf das mikrowellen-gestützte Abhören der Gespräche der Kriminellen erweitert werden kann.

* 8. Ausführungsbeispiel

* [0025]

Hirnforschung und Behandlung von Krankheiten. Die dargestellten Methoden der Gedankenübertragung ermöglichen neue Wege der Analyse, Therapie und Prophylaxe von bestimmten pathologischen Beeinträchtigungen des Hirnstoffwechsels und zur Beeinflussung von bestimmten nicht-pathologischen Limitierungen, Streßsituationen und Alterungsprozessen des Hirnstoffwechsels. Beispielsweise, da die elektromagnetische Strahlung in anderen Organteilen als bei der Anwendung von Schall oder sichtbarem Licht wirken kann, eröffnen sich neue Möglichkeiten. Z.B. bei Krankheiten können im Vergleich zu akustischen Reizen, die nicht auf der Einwirkung von elektromagnetischer Strahlung beruhen, andersartige Einwirkungen auf bestimmte neurologische Prozesse vorgenommen werden. Gedankenübertragung kann auch in der molekularen Medizin z.B. zur Analyse von biochemischen Netzwerken im Hirn unterstützend wirken. In einigen solchen Anwendungen kann es vorteilhaft sein, die Gedankenübertragung über wenige Millimeter Entfernung zu realisieren.

* 9. Ausführungsbeispiel

* [0026]

Unterstützung von Verhandlungen und Vorträgen von wichtigen Personen: Beispielsweise wird die Präsentation der wichtigen Person durch ein Team verfolgt, welches beratend mittels Gedankenübertragung eingreifen kann. An entscheidenden Stellen der Präsentation können z.B. wichtige Gedanken eingestreut werden. Bei unterschwelliger Gedankenübertragung wird der Vortragende – im Gegensatz zur herkömmlichen akustischen Übertragung mittels Ohrhörer – durch die Übertragung nicht gestört.

* 10. Ausführungsbeispiel

* [0027]

Beispiel für die Ermittlung der für die Erzeugung bestimmter Gedanken zu sendenden elektromagnetischen Signale in den Ausführungsbeispielen 1 – 9: Es werden Messungen einer großen Zahl von Korrelationen zwischen Stimuli und induzierten Gedanken oder Reaktionen durchgeführt. Diese Korrelationen werden mathematisch zusammengefaßt, um computergestützt Sequenzen von Stimuli erzeugen zu können, die besser mit gewünschten Gedanken oder Reaktionen korrelieren. Wenn z.B. 100 unabhängige Stimuli je eine 2%-ige Wahrscheinlichkeit

eines bestimmten Gedankens bewirken, können sie kombiniert eine ca. 87%-ige Wahrscheinlichkeit eines bestimmten Gedankens bewirken. Da viele der dargelegten Methoden der Gedankenübertragung über Monate anwendbar sind, ist es in vielen Fällen praktikabel, relativ schwache Korrelationen zu nutzen, um ein signifikantes Ergebnis zu erhalten.

* Abbildungsbeschreibungen

* [0028]

1 Relative Intensität (1) als Funktion der Frequenz in GHz (2), die unter bestimmten experimentellen Bedingungen notwendig ist, um impulsmodulierte Mikrowellenenergie akustisch wahrzunehmen. (nach Daten aus Lin, J.C. (1978) Microwave Auditory Effects and Applications. Charles C. Thomas, Publisher, Springfield, IL, USA). Bei hohen Frequenzen sinkt die Eindringtiefe in den Kopf, was zur Empfindlichkeitsverringerng führen kann (Stand der Technik).

* [0029]

2 Beispiel für einen Ausschnitt aus einem hochfrequenten amplitudenmodulierten Trägersignal. Die hochfrequente Strahlung, z.B. im Bereich von 1 – 1000 GHz, kann scharf gebündelt werden und breitet sich nahezu geradlinig aus. Die einhüllende Kurve des dargestellten Signals entspricht einem niederfrequenten Nutzsignal (z.B. 0,1 Hz – 1 MHz), welches z.B. in der Großhirnrinde, im Innenohr oder in anderen Organen wirksam wird.

* [0030]

3 Gedankenübertragung an einen Empfänger (3) mittels eines modulierten Strahls von Millimeterwellen oder Mikrowellen (4), der von einem MASER (5), z.B. Freie-Elektronen-MASER, ausgeht, welches auf ein Erkundungsfahrzeug (6) montiert ist, z.B. mittels eines Stativs (7). Der MASER kann z.B. ein Freie-Elektronen-MASER sein (auch oft als Freie-Elektronen-LASER bezeichnet). In einem Modus werden Sprachsignale eines Beobachters, z.B. mittels eines Mikrophons (8) eingegeben, im Erkundungsfahrzeug dem MASER-Strahl direkt amplitudenmoduliert. Zusätzlich können bewußtseinsmodifizierende Signale aufmoduliert werden. Solche bewußtseinsmodifizierende Signale sind z.B. hörbare Geräusche, die bestimmte Reaktionen bewußt auslösen können, oder unterschwellige Geräusche hörbarer Frequenz, die bestimmte Reaktionen unbewußt auslösen können, oder niederfrequente nicht-hörbare Signale (auf den elektromagnetischen Strahl modulierter Infraschall). Der Beobachter kann z.B. mittels des Detektors (9), z.B. ein Millimeterwellenteleskop oder ein Radardetektor, den Strahl nachführen und die Reaktion des Empfängers (3) verfolgen. Die Kombination von Detektor (9) und Computer (Computer mit ADC-Karte, Verstärker und Akkumulator (10); Anzeige (11); Keyboard (12); Joystick (13); Floppy Disc Drive (14); Schalter (15)) regelt vollautomatisch je nach Entfernungsunterschieden und absorbierenden Wänden, Bäumen oder Erdwällen die Intensität nach. Der Computer ist z.B. mittels eines Kabels (Kabel zur Verbindung zur Stromversorgung (16)) an eine Stromversorgung und z.B. mittels eines Kabelbündels (17) an eine Schrittmotormechanik (Gelenk und Schrittmotore (18)) zur Strahlnachführung angeschlossen. Je nach Wahl von Intensität des elektromagnetischen Strahls (4), dessen Modulation und Art der Einwirkung ist die Gedankenübertragung für den Empfänger (3) unbewußt oder bewußt. Die Reichweite der Gedankenübertragung beträgt z. B. 5 m – 20 km (19).

* [0031]

4 Beispiel für ein neuronales Netz (20) zur Berechnung der Sätze von Signalen (Stimuli (21)), die zur Erzeugung bestimmter Gedanken (Reaktionen (22)) gesendet werden. Die Sätze von Signalen

sind über neuronale Knoten, die bestimmten Übertragungsfunktionen entsprechen, mit den zu sendenden Gedanken (Reaktionen (22)) verbunden.

* [0032]

5 Handgetragenes Gedankenübertragungsgerät, das einen MASER (5), ein Mikrofon (8) zur Eingabe der Sprachsignale durch den Beobachter (23), eine aufladbare Energiequelle (Akkumulator (24)) und zur Beobachtung einen Detektor (9), z.B. eine Millimeterwellenkamera, enthält. Das Gedankenübertragungsgerät kann an das Stromfestnetz, das Stromnetz eines Fahrzeugs oder einen Generator (Stromgenerator (25)) mit z.B. 200 W Leistung angeschlossen werden. Mittels der Anzeige (11) und des Handgriffs (26) wird der elektromagnetische Strahl (4) des Gedankenübertragungsgeräts zum Empfänger (3) nachgeführt. Verschiedene Schalter (15) und die Elektronik (27) erlauben die Einstellung verschiedener Modi wie z.B. Sendung eingespeicherter Signale, automatische

Intensitätsanpassung, Art der Modulation zur Übertragung der Sprachsignale des Beobachters. Das Gedankenübertragungsgerät kann mittels eines Verbindungselements (28) auf Stative oder Fahrzeuge beweglich montiert werden. Die Reichweite der Gedankenübertragung und Beobachtung beträgt z.B. 5 m – 5000 m (29).

* [0033]

6 Gedankenübertragung an einen Empfänger (3) mittels eines modulierten Strahls von Millimeterwellen oder Mikrowellen (4), der von einem Phased Array (30) ausgeht, welches auf ein Erkundungsfahrzeug (6) montiert ist. Beispielsweise werden Worte computergespeichert und vom Computer in Impulsfolgen umgeformt, deren Einhüllende dem Intensitätsverlauf der Worte entspricht und dann die Impulsfolgen dem elektromagnetischen Strahl aufmoduliert und mit so geringen Intensitäten gesendet, daß der Empfänger (3) die Übertragung nicht bewußt wahrnimmt. Die Strahlnachführung erfolgt z.B. unter Nutzung der vom Empfänger (3) reflektierten Strahlung nach dem Radarprinzip. Die Reichweite der Gedankenübertragung beträgt z.B. 10 m – 1000 m (31).

* [0034]

7 Gedankenübertragung an einen Empfänger (3) mittels des Strahls (4) eines Phased Array (30) durch eine Stahlbetonwand (32) bei gleichzeitiger Beobachtung des Empfängers (3) mittels Millimeterwellenkamera (33). Stahlmaschen und kleinere metallische Gegenstände im Strahlengang stellen aufgrund der konischen Geometrie des Strahls kein erhebliches Problem dar. Gedankenübertragungsgerät und Millimeterwellenkamera sind z.B. auf einen Turm (34) montiert. Die Reichweite der Gedankenübertragung und Beobachtung beträgt z.B. 50 m – 5 km (35).

* [0035]

8 Gedankenübertragung an einen Empfänger (3) mittels des Strahls (4) eines in einem Gebäude (36) montierten MASERs (5) bei gleichzeitiger Beobachtung des Empfängers (3) mittels eines Detektors (Kamera (37)), z.B. Millimeterwellenkamera oder Infrarotkamera oder Detektor für die vom Empfänger (3) reflektierte Maserstrahlung. Die Nachführung des elektromagnetischen Strahls zum Empfänger (3) erfolgt computergesteuert (PC (38)). Zur Vermeidung unbeabsichtigter Nebenwirkungen ist die elektromagnetische Emission der Elektronik abgeschirmt (Abschirmung (39)). Zur Verbesserung der Reichweite kann sich das Gebäude z.B. auf einem Berg befinden. Die Reichweite der Gedankenübertragung und Beobachtung (zum Teil durch Gebäude (Gebäude einer Stadt (40)) und Wälder (Baum (41)) hindurch) beträgt z.B. 10 m – 200 km (42).

* [0036]

9 Gedankenübertragung von einem bemannten Flugzeug, einer unbemannten Drohne oder einem Helikopter (43) an einen Empfänger (3) mittels des speziell modulierten Strahls (4) eines MASERs (5). Die Reichweite der Gedankenübertragung beträgt z.B. 100 m – 20 km (44).

* [0037]

10 Gedankenübertragung von einem Satelliten (45) an Empfänger auf der Erde (46) mittels des Strahls (4) eines MASERs (5). Der MASER mit sehr kleinem Strahlöffnungswinkel wird von einer gepufferten starken Energiequelle, z.B. einer Kombination von Atombatterie und Akkumulator, gespeist. Zur Reduktion des Strahldurchmessers werden auch Selbstfokussierungseffekte der MASER-Strahlung genutzt. Die Reichweite der Gedankenübertragung beträgt z.B. 300 km – 800 km (47).

* [0038]

11 Gedankenübertragung an einige 100 bedeutende Empfänger (3) im Katastrophenfall mittels eines speziell modulierten elektromagnetischen Strahls (4). Zur besseren Detektion und Einstellung des elektromagnetischen Strahls (4) tragen die Empfänger (3) ein elektronisches Label. Die Übertragung erfolgt nach dem Multiplexprinzip quasisimultan durch schnelle Umschaltung der drei Phased Arrays (30) mit je 5000 W durchschnittlicher Sendeleistung. Die Reichweite der Gedankenübertragung (zum Teil durch Gebäude (40) hindurch) beträgt z.B. 50 m – 20 km (48).

* [0039]

12 Gedankenübertragung an Empfänger (3) im Notfall mittels eines modifizierten elektromagnetischen Gewehrs (Gewehr mit Teleskop (49)) zur Beobachtung und Betäubung von Empfängern (3) durch die Wände eines Gebäudes (Wand eines Gebäudes (50)) hindurch. Das Gewehr ist so modifiziert, daß es auch Gedanken mit geringer elektromagnetischer Strahlungsleistung übertragen und durch Wände hindurch hören kann (z.B. Detektion der Änderungen des Lungenvolumens).

KLASSIFIZIERUNGEN

Internationale Klassifikation A61M21/00

Unternehmensklassifikation A61M21/00, A61M2021/0055

Europäische Klassifikation A61M21/00

Patente zur elektromagnetischen Manipulation

Recherche „google“ Suchbegriff: patent number

google Roh-Übersetzung Englisch-Deutsch aus Programm der Suchmaschine

Formulierungsfehler durch elektronische, automatische Übersetzung!

Beispiel: US-Patent Nr. 3951134

Big Brother.....

Der Computer steuert einen zusätzlichen Übermittler, der das ausgleichende Signal dem Gehirn des Themas über eine Antenne übermittelt. Der Übermittler ist von der Hochfrequenzart, die allgemein in den Radaranwendungen benutzt wird. (Mobilfunkfrequenzband!)

Die Antenne.... ist ähnlich den Antennen.... und kann mit Ihnen kombiniert werden. Durch diese Mittel kann die Gehirnwellentätigkeit geändert werden und Abweichungen von einer gewünschten Norm können ausgeglichen werden. Gehirnwellen können überwacht werden und Steuersignale gesetzt werden, die dem Gehirn von einer entfernten Station übermittelt werden. Es soll angemerkt werden, dass die beschriebene Konfiguration eine der vielen Möglichkeiten ist, die formuliert werden können, ohne vom Geist meiner Erfindung abzuweichen.

Die Übermittler können monostatic und bistatic sein. Sie können einzelne, doppelte und mehrfache Frequenzvorrichtungen sein. Das übertragende Signal kann eine ununterbrochene Welle, Impuls FM oder jede mögliche Kombination von diesen sowie anderen Betriebsformen sein. Typische Arbeitsfrequenzen für die Übermittlung reichen von 1 MHz bis 40 Gigahertz, können jedoch geändert werden, um einer bestimmten Funktion, die überwacht werden und den Eigenschaften der spezifischen Anwendung zu entsprechen.

Die einzelnen Bestandteile des Systems für überwachend steuernde Gehirnwellentätigkeit können der herkömmlichen Art sein, die allgemein in den Radarsystemen eingesetzt wird (Mobilfunkfrequenzband!) Verschiedene Unterbaugruppen der Gehirnwellenüberwachung- und Kontrollapparate können hinzugefügt oder kombiniert werden. So können unterschiedliche Antennen oder einzelne multi-Modus-Antennen zum Senden oder Empfang benutzt werden. Zusätzliche Anzeigergeräte oder Computer können hinzugefügt werden, um den Geisteszustand festzustellen und die Gedankenprozesse zu überwachen.

Die Modulation des Störsignals, das durch das Gehirn noch mal übertragen wird, kann vom Umfang, von der Frequenz und/oder von der Phase sein. Passende Demodulatoren können benutzt werden, um die Gehirntätigkeit zu dechiffrieren und ausgewählte Bereiche der Gehirnwellen können durch den Computer analysiert werden, um den Geisteszustand festzustellen und die Gedankenprozesse zu überwachen. Apparate und Methode haben zahlreiche Anwendungsmöglichkeiten. Personen in kritischen Positionen wie Piloten können ununterbrochen überwacht werden. Körperliche Funktionen wie Pulsschlag, Herzrhythmus und andere können auch überwacht und das Auftreten von Halluzinationen festgestellt werden.....

Elektro-Forschung Elektromagnetische Spektrumeffekte auf biologische Systeme

„die Effekte der EM-Strahlungen auf biologische Systeme sind heute ein heißes Thema, da die Zahl künstlichen Quellen des EMF mehr und mehr zunehmen, mit einer konsequenten Notwendigkeit an die Sicherheitsstandard- und - risikobeurteilung. Das Verständnis der grundlegenden Einheiten und der Aufstellungsorte, der Abhängigkeit zwischen Feldern und Systemen ist kompliziert und verdient noch starke Untersuchungsbemühungen. Forschungsbemühungen werden hauptsächlich auf die Grundprinzipien der Verbindung zwischen EM-Energie und molekulare biologische Strukturen gerichtet. Mikroskopische Forschungsobjekte, die wert sind studiert zu werden, sind Proteinführungen innerhalb der Zellenmembrane. Jede Zellenmembrane wird durch die Proteinketten geführt und stellt mit einem Gatter versehen, stabilisierte gegenwärtige Flüsse durch Membranen dar. EM-Feldperturbe sind mit einem Gatter versehen und beeinflussen das physiologische Verhalten des Ganzen. Gegründet auf experimentelle mathematische Modelle der Beobachtungen entwickeln sich Techniken (Flecken-Klemmplattentechnik), um solche Phänomene zu erforschen, es sind interessante Resultate erreicht worden und es wurde geklärt wie und wo Effekte stattfinden. Wegen der sehr kleinen (fast Atom) Skala des beobachteten Systems und neuer Berechnungsannäherungen an diese Studien, basiert auf molekularer Dynamik, sind zur Zeit noch Entwicklung im Gange. Kommerzielle Pakete (GROMOS und MOPAC) werden besonders entwickelt, damit die bioelektromagnetische Abhängigkeit studiert werden kann. Bis jetzt sind Wasserlösungen, Peptid-Ketten und Ligandaufstellungsortkomplexbildung erfolgreich studiert worden.“, L. Tarricone

PATENTS FÜR ELECTROMAGNETICS UND BIOMANIPULATION

3951134 : Apparat und Methode um Gehirnwellen aus Entfernungen zu überwachen und ändern

4858612: Eine Methode und ein Apparat für Simulation des Hörens in den Säugetieren durch Einleitung einer Mehrzahl von Mikrowellen

3773049: APPARAT FÜR DIE BEHANDLUNG NEUROPSYCHOLOGISCHER UND KÖRPERLICHER KRANKHEITEN DURCH HITZE, LICHT, TON UND VHF 3576185: Schlafverursachung: METHODE UND ANORDNUNG MIT MODULIERTEM TON UND LICHT

5507291: Methode und Apparat für Informationen hinsichtlich des emotionalen Zustandes einer sich in Entfernung befindlichen Person festzustellen 4048986: Einzelne Kennzeichnung und Diagnose mit Wellenpolarisation

5458142: Vorrichtung für die Überwachung eines magnetischen Feldes, das von einem Organismus ausströmt 4951674: Analytisches biomagnetisches System mit magnetischen Glasfasersensoren

4591787: Mehrkanalvorrichtung mit KALMAREN und Superconducting Gradiometers für das Maß schwacher magnetischer Felder

4771239 : Mehrkanalvorrichtung mit Superconductor Gradiometers für das Messen der schwachen magnetischen Felder 3980076: Methode für die magnetischen Anfälligkeitsänderungen des menschlichen Körpers von außen

3789834: PROZESSE UND APPARAT FÜR DIE UNTERSUCHUNG DER INTERNEN PHYSIOLOGISCHEN PHÄNOMENE

5579241: Realzeiterwerbs- und archivierungssystem für mehrfache zeitgesteuerte Signale

5256960: Beweglicher Doppelstrahlungsmessapparat des elektromagnetischen Feldes des EMF-Bandes

5557199: Magnetresonanzmonitor

4864238: Vorrichtung für das Messen der schwachen magnetischen Flüsse, die planare Technologie verwenden 5330414: Gehirnwellen, die ein Apparat produziert

3884218: Methode des Verursachens und des Beibehaltens der verschiedenen Stadien des Schlafes im menschlichen Wesen

3837331: SYSTEM UND METHODE FÜR DAS STEuern DES NERVÖSEN SYSTEMS DES LEBENDEN ORGANISMUS

5213562: Methode des Verursachens der Geistes-, emotionalen und körperlichen Zustände des Bewusstseins, einschließlich der spezifischen Geistestätigkeit

3712292: METHODE UND APPARAT FÜR DAS PRODUZIEREN DER GEFEGTEN FM-AUDIO-Signal-Muster FÜR DAS VERURSACHEN DES SCHLAFES

4335710: Vorrichtung für die Induktion der spezifischen Gehirnwellenmuster

4573449: Methode zur Anregung des in Schlaf fallenden und/oder entspannenden Verhaltens eine Person und einer Anordnung

3835833: METHODE FÜR DAS ERHALTEN DER NEUROPHYSIOLOGISCHEN EFFEKTE
3727616: ELEKTRONISCHES SYSTEM FÜR DIE ANREGUNG DER BIOLOGISCHEN SYSTEME

3646940: VERPFLANZBARE ELEKTRONISCHE ANREGER-Elektrode UND -METHODE

3662758: ANREGER-Apparat FÜR MUSKULÖSE ORGANE MIT EXTERNEM ÜBERMITTLER UND VERPFLANZBARER EMPFÄNGER

4834701: Apparat für das Verursachen von Frequenzverringern der Gehirnwellen

5036858: Methode und Apparat für sich ändernde Gehirnwellenfrequenzen

4883067: Methode und Apparat für das Übersetzen des EEG in Musik, um psychologische und physiologische Zustände

4335710 zu verursachen und zu steuern: Vorrichtung für die Induktion der spezifischen Gehirnwellenmuster

4354505: Methode und Apparat für die Prüfung und die Anzeige des Entspannungszustandes eines menschlichen Themas

3967616: Mehrkanalsystem für und eine multifaktoriale Methode des Steuerns des nervösen Systems eines lebenden Organismus

5356368: Methode und Apparat für das Verursachen der gewünschten Zustände des Bewusstseins

5522386: Apparat besonders für Gebrauch in der Ermittlung des Zustandes des vegetativen Teils von nervösen Systemen

3893450: Methode und Apparat für Gehirnwellenformprüfung

5453361: Methode für das Produzieren des vom biologisch aktiven menschlichen Gehirn abgeleiteten neurotrophen Faktors

5124146: Differentielle Anlieferung der therapeutischen Mittel über der Blutgehirnsperre

4479932: Gehirn-spezifische Drogenanlieferung

4202323: Drogeaktivierung durch Strahlung

Links:

* Netz Des Geistigen Eigentums Delphion Such- und Forschungspatente durch Nummer und Schlüsselwort oder Konzept

* Akustische Technologie und das Gehirn - Patentsuche

* US-Patent- und -warenzeichenbüro

Patents for Electromagnetics and Biomanipulation

<http://www.catalase.com/patent.htm> <http://myweb.cableone.net/mtilton/trifaxpatents.html>
<http://tinyurl.com/c2n2s>

http://www.fraktali.biz/chemtrail/pdf/us_patents.pdf

US 2004013 8578A1

(19) United States

(12) Patent Application Publication

(10) Pub. NO.: us 2004/0138578 A1

(12) Пинеда и др.

(43) Дата публикации: Jul. 15,2004

(54) METHOD AND SYSTEM FOR A REAL TIME ADAPTIVE SYSTEM FOR EFFECTING CHANGES IN COGNITIVE-EMOTIVE PROFILES

(54) МЕТОД И СИСТЕМА ДЛЯ АДАПТИВНОЙ СИСТЕМЫ РЕАЛЬНОГО ВРЕМЕНИ ДЛЯ ОСУЩЕСТВЛЕНИЯ ИЗМЕНЕНИЯ В ПОЗНАВАТЕЛЬНО - ЭМОЦИОНАЛЬНОМ ПРОФИЛЕ (60)

Provisional application No. 60/398,508, filed on Jul. 25, 2002. (60)

Предварительная заявка № 60/398,508, зарегистрированная 25 июля 2002 г.

Publication Classification Классификация Публикации (51)

Int. Cl.7.....A61B 5/04 (52) U.S.

Cl.....600/544

US Patents: Electromagnetic Weapons

http://www.geocities.com/united_states_of_america_citizen/uspatents.htm

Omega - 13. Okt, 08:37

<http://www.rexresearch.com/sublimin/sublimin.htm>

© 2018 Strahlenfolter und Terror in Europa