



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 600 14 031 T2** 2005.01.20

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 141 943 B1**

(51) Int Cl.⁷: **G10L 17/00**

(21) Deutsches Aktenzeichen: **600 14 031.8**

(86) PCT-Aktenzeichen: **PCT/JP00/07238**

(96) Europäisches Aktenzeichen: **00 969 859.8**

(87) PCT-Veröffentlichungs-Nr.: **WO 01/029824**

(86) PCT-Anmeldetag: **18.10.2000**

(87) Veröffentlichungstag
der PCT-Anmeldung: **26.04.2001**

(97) Erstveröffentlichung durch das EPA: **10.10.2001**

(97) Veröffentlichungstag
der Patenterteilung beim EPA: **22.09.2004**

(47) Veröffentlichungstag im Patentblatt: **20.01.2005**

(30) Unionspriorität:
29974599 21.10.1999 JP

(84) Benannte Vertragsstaaten:
DE, FI, FR, GB, IT, SE

(73) Patentinhaber:
Casio Computer Co., Ltd., Tokio/Tokyo, JP

(72) Erfinder:
**SATOH, Katsuhiko, Hachioji-shi, JP; TAKEDA,
Tsuneharu, Hamura-shi, JP**

(74) Vertreter:
**Grünecker, Kinkeldey, Stockmair &
Schwanhäusser, 80538 München**

(54) Bezeichnung: **SPRACHERERKENNUNG DURCH KORRELIERUNG VON SPEKTROGRAMMEN**

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

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

Beschreibung

Technisches Gebiet

[0001] Die vorliegende Erfindung bezieht sich auf eine Sprachvergleichsvorrichtung und ein Sprachvergleichsverfahren für das Identifizieren einer Person mit Sprachdaten.

Stand der Technik

[0002] Im Allgemeinen wird für das Identifizieren eines Sprechers mit einer Sprache ein zu vergleichendes Sprachsignal vor einem Vergleich in einen akustischen Parameter wie das Frequenzspektrum oder Ähnliches umgewandelt, da es nicht effizient ist, das Sprachsignal direkt mit registrierten Sprachsignalen zu vergleichen. Andere für den Zweck verfügbare Parameter können die Grundfrequenz ("Principle"-Frequenz, "Pitch"-Frequenz), Sprachenergie, Formantenfrequenz, Nullstellenzahl und Ähnliches sein.

[0003] Da diese akustischen Parameter in erster Linie phonetische Information und personengebundene Information in zweiter Linie beinhalten, muss hier für einen Vergleich aus den akustischen Parametern eine neue charakteristische sprechereindeutige Größe ("amount") erzeugt werden, um die Trefferquote zu verbessern, wenn ein Sprecher identifiziert wird.

[0004] Eine konventionelle Sprecheridentifikation wird auf die folgende Art und Weise durchgeführt.

[0005] Fig. 14 ist ein Flussdiagramm, das einen Ablauf einer Sprecheridentifikation mittels einer konventionellen Sprachvergleichsvorrichtung illustriert.

- (1) Ein für ein Wort geäußertes Eingabesprachsignal wird in Rahmen von vorgegebener Zeiteinheit unterteilt, und für jeden der Rahmen wird das Frequenzspektrum berechnet, um eine Zeitreihenverteilung der Frequenzspektren (im Folgenden bezeichnet als "Soundspektrogramm") abzuleiten (Schritt C1).
- (2) Ein Sprachabschnitt wird aus dem Soundspektrogramm ermittelt (Schritt C2).
- (3) Es wird bestimmt, ob der Sprachabschnitt ein gesprochener, ein nicht gesprochener oder ein stiller Abschnitt ist, um die gesprochenen Abschnitte aus dem Sprachabschnitt zu extrahieren. Dann wird der Sprachabschnitt in Blöcke unterteilt, von denen jeder jedem der gesprochenen Abschnitte entspricht (Schritt C3).
- (4) Als charakteristische sprechereindeutige Größe wird für die Blöcke ein additiver Mittelwert des Soundspektrogramms in der Zeitrichtung (im Folgenden bezeichnet als "Durchschnittsspektrum") berechnet (Schritt C4).
- (5) Es wird bestimmt, ob die Verarbeitung der Registrierung oder dem Vergleich dient, und das Durchschnittsspektrum für die Blöcke wird als charakteristische Größe eines registrierten Sprechers registriert, wenn die Registrierung vorgesehen ist (Schritte C5 → C6).
- (6) Es wird bestimmt, ob die Verarbeitung der Registrierung oder dem Vergleich dient, und die Ähnlichkeit bezüglich der charakteristischen Größe des registrierten Sprechers wird mit dem Durchschnittsspektrum der Blöcke, welches als charakteristische Größe eines unbekannten Sprechers verwendet wird, berechnet (Schritt C5 → C7).
- (7) Die Ähnlichkeit des unbekannten Sprechers mit dem registrierten Sprecher wird mit einem vorher gesetzten Schwellwert verglichen, um die Identität des registrierten Sprechers mit dem unbekannten Sprecher zu bestimmen (Schritt C8).

[0006] Wie oben beschrieben, vergleicht der durch die konventionelle Sprachvergleichsvorrichtung durchgeführte Sprecheridentifikationsablauf ein von einem registrierten Sprecher eingegebenes Sprachsignal (im Folgenden bezeichnet als das "registrierte Sprachsignal") mit einem von einem unbekannten Sprecher für den Vergleich eingegebenen Sprachsignal (im Folgenden bezeichnet als das "unbekannte Sprachsignal"), indem er (1) das Sprachsignal in das Soundspektrogramm konvertiert; (2) einen Sprachabschnitt aus dem Soundspektrogramm ermittelt; (3) einen gesprochenen Abschnitt aus dem ermittelten Sprachabschnitt extrahiert auf Basis einer Bestimmung, ob der Sprachabschnitt ein gesprochener, nicht gesprochener oder stiller Abschnitt ist; und (4) eine charakteristische Größe für jeden von aus dem extrahierten gesprochenen Abschnitt unterteilten Blöcken ableitet. Auf diese Weise bringt die Berechnung der charakteristischen Größe, die auf die Vergleichsbearbeitung angewandt wird, um tatsächlich ("actually") die Identität des registrierten Sprachsignals mit dem unbekannten Sprachsignal zu bestimmen, mindestens vier Vorverarbeitungsstufen mit sich, so dass eine große Zahl von Verarbeitungsschritten für die gesamte Sprecheridentifikationsverarbeitung benötigt wird.

[0007] Ferner benötigt, obwohl der konventionelle Sprecheridentifikationsablauf, der das additive Mittel des Soundspektrogramms in einem Block in der Zeitrichtung als eine charakteristische sprechereindeutige Größe verwendet, in seiner relativ einfachen Verarbeitung vorteilhaft ist, die Erzeugung einer stabilen charakteristischen Größe Sprachsignalen für eine relativ lange Zeitspanne. Zusätzlich ist dieser Ablauf nicht für textabhängige Sprecheridentifikation geeignet, da Information in der Zeitachsenrichtung komprimiert wird. Weiterhin wird keine hinreichende charakteristische Größe bereitgestellt, da der konventionelle Sprecheridentifikationsablauf als Begleiterscheinung der Mittelung der phonetischen Information personengebundene Information mittelt, die der phonetischen Information überlagert ist. Daher muss, um die Trefferquote zu verbessern, eine zusätzliche charakteristische Größe hinzugefügt werden, was zur Folge hat, dass eine extrem große Anzahl von Vorverarbeitungsschritten benötigt wird.

[0008] Daher impliziert die Verbesserung der Trefferquote das Problem einer damit einhergehenden extrem großen Anzahl von Vorverarbeitungsschritten.

Offenbarung der Erfindung

[0009] Entsprechend ist es die Aufgabe der vorliegenden Erfindung, eine Sprachvergleichsvorrichtung und ein Sprachvergleichsverfahren bereitzustellen, welche fähig sind, einen Sprecher mit einer hohen Trefferquote ohne das Erfordernis einer großen Anzahl von Vorverarbeitungsschritten zu identifizieren.

[0010] Die vorliegende Erfindung wird durch die unabhängigen Ansprüche definiert. Ausgestaltungen der Erfindung werden durch die abhängigen Ansprüche definiert.

[0011] Entsprechend der vorliegenden Erfindung wird eine Sprachvergleichsvorrichtung bereitgestellt, welche Datenkonvertierungsmittel für die Konvertierung von zwei einem Vergleich unterzogenen Sprachsignalen in zwei zweidimensionale Daten, die kennzeichnend für Sprachcharakteristika der zwei Sprachsignale sind; Schablonen ("templates")-Platzierungsmittel für das Platzieren einer Vielzahl von Schablonen für das Definieren einer Vielzahl von Teilflächen ("areas") auf einen der zweidimensionalen Daten; Korrelierte-Teilflächen-Ermittlungsmittel für das Ermitteln von Teilflächen auf den anderen der zweidimensionalen Daten, die eine maximale Korrelation bezüglich einer Vielzahl von Teilflächen auf den anderen der zweidimensionalen Daten haben und der Vielzahl von Schablonen entsprechen; und Vergleichsbestimmungsmittel für das Vergleichen eines gegenseitigen Lageverhältnisses der Vielzahl von Schablonen auf den einen der zweidimensionalen Daten mit einem gegenseitigen Lageverhältnis der Vielzahl von Teilflächen auf den anderen der zweidimensionalen Daten, die durch das Korrelierte-Teilflächen-Ermittlungsmittel ermittelt wurden, um Identität zwischen den beiden Sprachsignalen zu bestimmen, umfasst.

[0012] Zusätzliche Vorteile der vorliegenden Erfindung werden in der folgenden Beschreibung dargelegt werden und werden teilweise aus der Beschreibung naheliegend sein oder können aus der Praxis der vorliegenden Erfindung erfahren werden.

[0013] Die Aufgabe und Vorteile der vorliegenden Erfindung können mittels der im Folgenden genau aufgezeigten Vermittlungen und Kombinationen realisiert und erhalten werden.

Kurze Beschreibung der Zeichnungen

[0014] Die beigefügten Zeichnungen, welche in die Beschreibung eingefügt sind und einen Teil derselben bilden, illustrieren gegenwärtig bevorzugte Ausgestaltungen der vorliegenden Erfindung und dienen zusammen mit der obigen allgemeinen Beschreibung und der nachstehend angeführten genauen Beschreibung der bevorzugten Ausgestaltungen dazu, die Prinzipien der vorliegenden Erfindung zu erklären, in denen:

[0015] **Fig. 1** ein Blockdiagramm ist, das die Konfiguration eines elektrischen Schaltkreises in einer Sprachvergleichsvorrichtung entsprechend einer ersten Ausgestaltung der vorliegenden Erfindung illustriert;

[0016] **Fig. 2** ein Blockdiagramm ist, das die Konfiguration von operationellen Funktionen in der Ausführung eines Sprechregistrierungs-/vergleichsverarbeitungsprogramms in der ersten Ausgestaltung der Sprachvergleichsvorrichtung illustriert;

[0017] **Fig. 3** ein Flussdiagramm ist, das die Sprachregistrierungs-/vergleichsverarbeitung in der ersten Ausgestaltung der Sprachvergleichsvorrichtung illustriert;

[0018] Fig. 4A und 4B zeigen, wie Sprachsignalaten in Soundspektrogrammdateien in einer Soundspektrogramm-Konvertierungseinheit in der Sprachvergleichsvorrichtung konvertiert werden;

[0019] Fig. 5A und 5B einen Vergleich eines registrierten Sprachspektrogramms mit einem unbekannten Sprachspektrogramm und eine Platzierung einer Schablone auf dem registrierten Sprachspektrogramm entsprechend der Sprachregistrierungs-/vergleichtsverarbeitung in der Sprachvergleichsvorrichtung zeigen;

[0020] Fig. 6A und 6B einen Vergleich einer Verteilung von Positionen, an denen Schablonenteilflächen auf einem registrierten Sprachspektrogramm platziert sind, mit einer Verteilung von Positionen, an denen maximal korrelierte Teilflächen auf einem unbekannten Sprachspektrogramm ermittelt wurden, entsprechend der Sprachregistrierungs-/vergleichtsverarbeitung in der Sprachvergleichsvorrichtung zeigt;

[0021] Fig. 7 ein Blockdiagramm ist, das die Konfiguration von operationellen Funktionen in der Ausführung eines Sprachregistrierungs-/vergleichtsverarbeitungsprogramms entsprechend einer zweiten Ausgestaltung der Sprachvergleichsvorrichtung der vorliegenden Erfindung illustriert;

[0022] Fig. 8 ein Flussdiagramm ist, das die Sprachregistrierungs-/vergleichtsverarbeitung in der zweiten Ausgestaltung der Sprachvergleichsvorrichtung illustriert;

[0023] Fig. 9 ein Blockdiagramm ist, das die Konfiguration von operationellen Funktionen in der Ausführung eines Sprachregistrierungs-/vergleichtsverarbeitungsprogramms entsprechend einer dritten Ausgestaltung der Sprachvergleichsvorrichtung der vorliegenden Erfindung illustriert;

[0024] Fig. 10 ein Flussdiagramm ist, das die Sprachregistrierungs-/vergleichtsverarbeitung in der dritten Ausgestaltung der Sprachvergleichsvorrichtung illustriert;

[0025] Fig. 11 zeigt, wie ein Sprachabschnitt aus einem Soundspektrogramm "A" einer registrierten Sprache in einer Sprachabschnittsermittlungseinheit in der Sprachvergleichsvorrichtung der dritten Ausgestaltung ermittelt wird;

[0026] Fig. 12A und 12B einen Vergleich eines registrierten Sprachspektrogramms mit einem unbekannten Sprachspektrogramm und eine Platzierung einer Schablone auf dem registrierten Sprachspektrogramm in der Sprachregistrierungs-/vergleichtsverarbeitung in der Sprachvergleichsvorrichtung der dritten Ausgestaltung zeigen;

[0027] Fig. 13A und 13B einen Vergleich einer Verteilung von Positionen, an denen Schablonenteilflächen auf einem registrierten Sprachspektrogramm platziert sind, mit einer Verteilung von Positionen, an denen maximal korrelierte Teilflächen auf einem unbekannten Sprachspektrum ermittelt wurden, entsprechend der Sprachregistrierungs-/vergleichtsverarbeitung in der Sprachvergleichsvorrichtung der dritten Ausgestaltung zeigen; und

[0028] Fig. 14 ein Flussdiagramm ist, das einen Sprecheridentifikationsablauf illustriert, der durch eine konventionelle Sprachvergleichsvorrichtung durchgeführt wird.

Beste Art und Weise für die Ausführung der Erfindung

[0029] Eine bevorzugte Ausgestaltung einer Sprachvergleichsvorrichtung und eines Sprachvergleichsverfahrens entsprechend der vorliegenden Erfindung wird nun bezugnehmend auf die beigefügten Zeichnungen beschrieben werden.

Erste Ausgestaltung

[0030] Fig. 1 ist ein Blockdiagramm, das die Konfiguration eines elektrischen Schaltkreises in einer Sprachvergleichsvorrichtung entsprechend der ersten Ausgestaltung der vorliegenden Erfindung illustriert.

[0031] Die illustrierte Sprachvergleichsvorrichtung umfasst einen Computer-basierten Controller (CPU) **11**. Die CPU **11** reagiert auf ein Eingabesignal von einer Eingabeeinheit **12**, um ein zuvor in einem Speicherelement **13** gespeichertes Systemprogramm; ein von einem externen Aufzeichnungsmedium **14** wie einer Diskette oder Ähnlichem über ein Diskettenlaufwerk in das Speicherelement **13** gelesenes Sprachregistrierungs-/vergleichtsverarbeitungsprogramm; oder ein von einem externen Computer über ein Kommunikations-

netzwerk wie das Internet und einen Übertragungscontroller **15** in das Speicherelement **13** heruntergeladenes Sprachregistrierungs-/vergleichsverarbeitungsprogramm zu initiieren, und kontrolliert den Betrieb von Komponenten der Vorrichtung mit einer RAM **16**, die als Arbeitsspeicher benutzt wird.

[0032] Die CPU **11** ist zusätzlich zur Eingabeeinheit **12**, dem Speicherelement **13**, dem Übertragungscontroller **15** und der RAM **16** mit einer Ausgabeeinheit **17** und einem Drucker **18** verbunden.

[0033] Die Eingabeeinheit **12** umfasst eine Tastatur, eine Maus und ein Mikrofon für das Eingeben einer Sprache.

[0034] Das Speicherelement **13** speichert das Sprachregistrierungs-/vergleichsverarbeitungsprogramm, welches Sprachdaten registriert, die für eine Sprache kennzeichnend sind, die von einem Nutzer für einen Satz oder ein Wort geäußert und in die Sprachvergleichsvorrichtung eingegeben wurde, und die registrierten Sprachdaten mit neu eingegebenen Sprachdaten vergleicht, die für denselben Satz oder dasselbe Wort kennzeichnend sind, um die Identität dazwischen zu bestimmen, und speichert ferner für die Registrierung zweidimensionale Daten, die aus zu registrierenden Sprachdaten konvertiert wurden, welche entsprechend der Sprachregistrierungs-/vergleichsverarbeitung eingegeben werden.

[0035] Die RAM **16** speichert temporär zweidimensionale Daten, die für ein Soundspektrum kennzeichnend sind, das aus Sprachdaten konvertiert wurde, welche entsprechend der Sprachregistrierungs-/vergleichsverarbeitung mit registrierten Sprachdaten verglichen werden sollten. Das Soundspektrum der in dem Speicherelement **13** gespeicherten registrierten Sprachdaten wird dem Soundspektrum der in der RAM **16** gespeicherten unbekannten Sprachdaten zugeordnet, um die Identität dazwischen zu bestimmen.

[0036] In diesem Fall wird der Vergleich des Soundspektrums der registrierten Sprache mit dem Soundspektrum der unbekannten Sprache durchgeführt, indem z. B. rechteckige Schablonen an einer Vielzahl von Positionen auf den zweidimensionalen Daten platziert werden, die die registrierte Sprache repräsentieren, eine Teilfläche ermittelt wird, an welcher auf den unbekannten zweidimensionalen Sprachdaten ein maximaler Korrelationskoeffizient für alle registrierten zweidimensionalen Sprachdaten innerhalb der Vielzahl von Schablonen erhalten wird, und eine Verteilung der Positionen, an denen die Schablonen auf den registrierten zweidimensionalen Sprachdaten platziert sind (gegenseitiges Lageverhältnis) mit einer Verteilung der Positionen, an denen der maximale Korrelationskoeffizient auf den unbekannten zweidimensionalen Sprachdaten ermittelt wird (gegenseitiges Lageverhältnis) verglichen wird.

[0037] Dann kann das Ergebnis der Bestimmung der Identität zwischen der registrierten Sprache und der unbekannten Sprache an der Ausgabeeinheit **17** ausgegeben werden oder durch den Drucker **18** gedruckt werden.

[0038] Fig. 2 ist ein Blockdiagramm, das die Konfiguration von operationellen Funktionen entsprechend der Ausführung des Sprachregistrierungs-/vergleichsverarbeitungsprogramms in der ersten Ausgestaltung der Sprachvergleichsvorrichtung illustriert.

[0039] In den funktionellen Blöcken, die kennzeichnend für die operationellen Funktionen in der Sprachvergleichsvorrichtung sind, sind entsprechende Komponenten in dem elektronischen Schaltkreis in Fig. 1 zusätzlich durch Bezugszeichen in Klammern bezeichnet.

[0040] Eine Sprachsignaldaten-Eingabeeinheit **21** verwendet ein Mikrofon oder Ähnliches, um zu registrierende Sprache oder zu vergleichende Sprache in ein elektrisches Signal zu konvertieren, und gibt das resultierende elektrische Signal in eine nachfolgende Einheit ein. Sprachsignaldaten, die kennzeichnend für eine registrierte Sprache oder eine unbekannte Sprache sind, die über die Sprachsignaldaten-Eingabeeinheit **21** eingegeben wurde, werden einer Sprachspektrum-Konvertierungseinheit **22** zugeführt.

[0041] Die Soundspektrum-Konvertierungseinheit **22** konvertiert Sprachsignaldaten, die für eine registrierte Sprache oder eine unbekannte Sprache kennzeichnend sind, die von der Sprachsignaldaten-Eingabeeinheit **21** eingegeben wurden in zweidimensionale Daten (Soundspektrum), welche eine Zeitserienverteilung des Frequenzspektrums in Form einer Dichte repräsentieren, so dass eine stärkere ("more") Frequenzkomponente heller erscheint und eine schwächere ("less") Frequenzkomponente dunkler erscheint auf einem Koordinatensystem, welches die Frequenz auf der vertikalen Achse und die Zeit auf der horizontalen Achse angibt (vgl. Fig. 4A und 4B). Das Soundspektrum der registrierten Sprache oder der unbekannten Sprache, das durch die Soundspektrum-Konvertierungseinheit **22** konvertiert wurde, wird für die registrierte

Sprache einer Registrierter-Sprecher-Informationseinheit **23** und für die unbekannte Sprache einer Maximal-Korrelierte-Teilflächen-Ermittlungseinheit **26** einer Vergleichseinheit **24** zugeführt.

[0042] Die Registrierter-Sprecher-Informationsspeichereinheit **23** registriert Daten, die kennzeichnend für registrierte Sprachen in Form von Soundspektren sind, die durch die Soundspektrogramm-Konvertierungseinheit **22** konvertiert wurden.

[0043] Eine Schablonenplatzierungseinheit **25** der Vergleichseinheit **24** definiert eine Vielzahl rechteckiger Teilflächen (Schablonen) von beliebiger Größe und platziert die Schablonen an beliebigen Positionen auf einem Soundspektrogramm einer registrierten Sprache, das in der Registrierter-Sprecher-Informationsspeichereinheit **23** gespeichert ist (vgl. **Fig. 5**). Zweidimensionale Daten innerhalb der Teilflächen, die jeweils der Vielzahl von Schablonen entsprechen, die durch die Schablonenplatzierungseinheit **25** platziert wurden, werden der Maximal-Korrelierte-Teilflächen-Ermittlungseinheit **26** zugeführt.

[0044] Die Maximal-Korrelierte-Teilflächen-Ermittlungseinheit **26** ermittelt eine Vielzahl von Teilflächen (maximal korrelierten Teilflächen), die für die zweidimensionalen Daten innerhalb der Teilflächen, die jeweils der Vielzahl von Schablonen auf dem registrierten Soundspektrogramm entsprechend, das von der Schablonenplatzierungseinheit **25** zugeführt wurde, maximale Korrelationen auf einem unbekannten Sprachsoundspektrogramm haben (vgl. **Fig. 6**). Koordinatendaten, die kennzeichnend sind für die jeweiligen Positionen der Vielzahl von Schablonen, die durch die Schablonenplatzierungseinheit auf dem registrierten Sprachsoundspektrogramm platziert wurden, und Koordinatendaten, die kennzeichnend sind für die jeweiligen Positionen einer Vielzahl von maximal korrelierten Teilflächen, die durch die Maximal-Korrelierte-Teilflächen-Ermittlungseinheit **26** ermittelt wurden, werden einer Vergleichsbestimmungseinheit **27** zugeführt.

[0045] Die Vergleichsbestimmungseinheit **27** stützt sich auf die Koordinatendaten, die kennzeichnend sind für die jeweiligen Positionen der Vielzahl von Schablonen auf dem registrierten Sprachsoundspektrogramm, das von der Schablonenplatzierungseinheit **25** zugeführt wurde und auf die Koordinatendaten, die kennzeichnend sind für die jeweiligen Positionen der Vielzahl von maximal korrelierten Teilflächen, die durch die Maximal-Korrelierte-Teilflächen-Ermittlungseinheit ermittelt wurden, um eine Verteilung der Positionen, an denen die Schablonen auf dem registrierten Sprachsoundspektrogramm platziert sind (gegenseitiges Lageverhältnis), mit einer Verteilung der Positionen, an denen die maximal korrelierten Teilflächen auf dem unbekannten Sprachsoundspektrogramm ermittelt wurden (gegenseitiges Lageverhältnis), zu vergleichen, um für eine Evaluierung der Identität die Ähnlichkeit der registrierten Sprache mit der unbekannten Sprache aus einem Unterschied zwischen denselben zu bestimmen. Das Ergebnis der Evaluierung für die Identität zwischen der registrierten Sprache und der unbekannten Sprache, das durch die Vergleichsbestimmungseinheit **27** produziert wurde, wird eine Vergleichsergebnis-Ausgabereinheit **28** zugeführt und dort ausgegeben.

[0046] Als Nächstes wird die Sprachvergleichsvorrichtung insbesondere hinsichtlich des Betriebs mit der oben beschriebenen Konfiguration beschrieben werden.

[0047] **Fig. 3** ist ein Flussdiagramm, das eine Sprachregistrierungs-/vergleichtsverarbeitung in der ersten Ausgestaltung der Sprachvergleichsvorrichtung illustriert.

[0048] **Fig. 4A** und **4B** zeigen, wie Sprachsignaldaten in der Soundspektrogramm-Konvertierungseinheit **22** in der Sprachvergleichsvorrichtung in Soundspektrogrammdaten konvertiert werden.

[0049] **Fig. 5A** und **5B** zeigen einen Vergleich eines registrierten Sprachspektrogramms mit einem unbekannten Sprachspektrogramm (zu vergleichenden Daten) und eine Platzierung von Schablonen auf dem registrierten Sprachspektrogramm entsprechend der Sprachregistrierungs-/vergleichtsverarbeitung in der Sprachvergleichsvorrichtung.

[0050] **Fig. 6A** und **6B** zeigen einen Vergleich einer Verteilung von Positionen, an denen Schablonenteilflächen auf einem registrierten Sprachspektrogramm platziert sind mit einer Verteilung von Positionen, an denen maximal korrelierte Teilflächen auf einem unbekannten Sprachspektrogramm (zu vergleichenden Daten) ermittelt wurden entsprechend der Sprachregistrierungs-/vergleichtsverarbeitung in der Sprachvergleichsvorrichtung.

[0051] Für die Verwendung der Sprachvergleichsvorrichtung als Personenidentifikationsvorrichtung werden zuerst Sprachsignaldaten einer registrierten Person, die der Identifikation unterzogen wird, von der Sprachsignaldaten-Eingabereinheit **21** eingegeben. Diese eingegebenen registrierten Sprachsignaldaten a(n) werden

wie in **Fig. 4A** und **4B** (Schritte S1 und S2) gezeigt in der Soundspektrogramm-Konvertierungseinheit **22** in Zeitseriendaten $A(n, f)$ eines Frequenzspektrums (im Folgenden bezeichnet als das Soundspektrogramm "A") konvertiert.

[0052] Die Konvertierungsverarbeitung, die in der Soundspektrogramm-Konvertierungseinheit **22** durchgeführt wird, um die Sprachsignaldaten in das Soundspektrogramm zu konvertieren, bringt eine Unterteilung der Sprachsignaldaten $d(n)$ in Rahmen in Intervallen von vorgegebener Zeiteinheit (die Anzahl von Abtastpunkten ist N), ein Multiplizieren der unterteilten Datenserie $d_N(n)$ mit einer Humming Fensterfunktion $w_N(n)$ als Fensterfunktion, ein Fast Fourier Transformieren (FFT) des Multiplikationsergebnisses, um Frequenzspektren $S(f)$ zu berechnen, ein Anordnen der für die jeweiligen unterteilten Rahmen berechneten Frequenzspektren $S(f)$ in eine Zeitreihe und ein Konvertieren der Frequenzspektren $S(f)$ in ein Soundspektrogramm $S(n, f)$, in welchem die horizontale Achse die Zeit und die vertikale Achse die Frequenz repräsentiert, mit sich. Es sollte angemerkt werden, dass ein Spektralwert abgeleitet wird, indem ein Logarithmus von einem Originalwert genommen wird, und der logarithmische Wert mit einem Maximalwert von 255 normalisiert wird. In anderen Worten hat das Soundspektrogramm das Aussehen ("aspect") zweidimensionaler Multilevelabstufungsdaten, d. h. Multilevelabstufungsbilddaten, so dass diese Ausgestaltung das Soundspektrogramm als Multilevelabstufungs-Stimmenabdrucksbild auf die Sprecheridentifikation anwendet.

[0053] Im Folgenden wird das Soundspektrogramm als Multilevelabstufungsbilddaten betrachtet, in denen jedes Wort Pixeldaten entspricht.

[0054] Das Soundspektrogramm "A", das einer Sprache eines registrierten Sprechers entspricht, die durch die Soundspektrogramm-Konvertierungseinheit **22** erhalten wurde, wird in der Registrierter-Sprecher-Informationsspeichereinheit **23** gespeichert und als eine charakteristische Größe des registrierten Sprechers registriert (Schritt S3).

[0055] Andererseits, da Sprachsignaldaten eines unbekannten Sprechers, der der Identifikation in der Personenidentifikationsvorrichtung unterzogen wird, über die Sprachsignaldaten-Eingabeeinheit **21** eingegeben werden, werden die eingegebenen unbekannten Sprachsignaldaten $b(n)$ in der Soundspektrogramm-Konvertierungseinheit **22** in ein Soundspektrogramm $B(n, f)$ (im Folgenden bezeichnet als das Spektrogramm "B") konvertiert auf eine Art und Weise ähnlich den Sprachdaten des registrierten Sprechers (Schritte S1 bis S4).

[0056] Anschließend wird in der Schablonenplatzierungseinheit **25** in der Vergleichseinheit **24** das Soundspektrogramm "A" der registrierten Sprache, das in der Registrierter-Sprecher-Informationsspeichereinheit **22** registriert ist, gelesen. Dann werden wie in **Fig. 5A** gezeigt eine Vielzahl von rechteckigen Teilflächen, d. h. Schablonen t_i ($i = 1$ bis m , wobei m eine ganze Zahl größer oder gleich zwei ist) definiert und im Soundspektrogramm "A" platziert (Schritt S5).

[0057] Anschließend rasterscannt die Maximal-Korrelierte-Teilflächen-Ermittlungseinheit **26** in der Vergleichseinheit **24** eine rechteckige Teilfläche, die jeder der Schablonen t_i entspricht, die im Soundspektrogramm "A" der registrierten Sprache definiert sind, auf einer Pixel für Pixel-Basis in der horizontalen Richtung und in der vertikalen Richtung auf dem Soundspektrogramm "B" der unbekannten Sprache, um sequenziell einen Korrelationskoeffizienten zu berechnen, wobei sie alle Pixeldaten in der zugehörigen Schablone t_i und Pixeldaten im Soundspektrogramm "B" der unbekannten Sprache in der rechteckigen Teilfläche benutzt. Somit wird, wie in **Fig. 6B** (Schritte S6 und S7) gezeigt, eine Teilfläche T_i im Soundspektrogramm "B" einer unbekannten Sprache ermittelt, in welcher ein maximaler Korrelationskoeffizient vorhanden ist. Wie der Korrelationskoeffizient berechnet wird, wird später beschrieben werden.

[0058] Die Berechnung des maximalen Korrelationskoeffizienten auf dem Soundspektrogramm "B" der unbekannten Sprache, die auf den Bilddaten jeder Schablone t_i , die im Soundspektrogramm "A" der registrierten Sprache definiert ist, basiert, und die Ermittlung der Teilfläche T_i werden sequenziell für jede Schablone t_i durchgeführt (Schritte S6 bis S8). Dann, nachdem bestimmt wurde, dass die jeweiligen Teilflächen T_i auf dem Soundspektrogramm "B" der unbekannten Sprache für alle Schablonen t_i als diejenigen ermittelt worden sind, die maximale Korrelationskoeffizienten haben, vergleicht die Vergleichsbestimmungseinheit **27** in der Vergleichseinheit **24**, wie in **Fig. 6A** und **6B** gezeigt, eine Verteilung der jeweiligen Schablonenteilflächen $\{t_i\}$, die im Soundspektrogramm "A" der registrierten Sprache platziert sind (Lageverhältnis), mit einer Verteilung der jeweiligen Teilflächen $\{T_i\}$, die aus dem Soundspektrogramm "B" der unbekannten Sprache ermittelt wurden (Lageverhältnis), basierend auf ihren jeweiligen Koordinatendaten, um die Identität zwischen denselben zu evaluieren (Schritte S8 und S9).

[0059] Im Folgenden wird ein spezielles Beispiel beschrieben werden.

[0060] Die Koordinaten, die als Basis dienen für ein Spezifizieren der Positionen einer Schablone t_1 , die auf dem Soundspektrum "A" der registrierten Sprache platziert ist, und einer rechteckigen Teilfläche T_1 auf dem Bild, das aus dem Soundspektrum "B" der unbekannten Sprache ermittelt wurde, z. B. die Koordinaten der linken oberen Ecken der beiden Rechtecke, werden erfasst ("acquired") und jeweils an vorgegebenen Stellen der RAM **16** als $t_1(X_1, Y_1)$ und $T_1(X_{T_1}, Y_{T_1})$ gespeichert.

[0061] Als Nächstes werden Schablonen t_2 bis t_5 im Soundspektrum "A" der registrierten Sprache definiert, so dass diese Schablonen auf den jeweiligen Ecken der Schablone t_1 zentriert sind. Es sollte hierbei angemerkt werden, dass die Platzierung der definierten Schablonen beliebig ist, und dass die Anzahl der Schablonen nicht auf vier beschränkt ist, sondern jede beliebige Anzahl sein kann, solange eine erforderliche Vergleichsgenauigkeit für den Sprachvergleich gewährleistet werden kann. Ferner kann auch die Größe der hierbei definierten Schablonen beliebig sein.

[0062] Dann, ähnlich der Ermittlung der rechteckigen Teilfläche T_1 , wird in jeder der Schablonen t_2 bis t_5 (im Folgenden repräsentiert durch t_i ($i = 2, 3, 4, 5$)) eine rechteckige Teilfläche, die dieselbe Größe hat wie die Schablone t_i , im Soundspektrum "B" der unbekannten Sprache gesetzt. Die rechteckige Teilfläche wird zweidimensional in Einheiten von Pixeln verschoben, und ein Korrelationskoeffizient zwischen der rechteckigen Teilfläche und der Schablone t_i wird jedes Mal berechnet, wenn die rechteckige Teilfläche verschoben wird. Schließlich wird eine rechteckige Teilfläche an einer Position, an welcher ein maximaler Korrelationskoeffizient vorhanden ist, als rechteckige Teilfläche T_i ausersehen ("designated"), und die Koordinaten, welche als Basis für ein Spezifizieren der Positionen der Schablone t_i und der rechteckigen Teilfläche T_i auf dem Bild dienen, z. B. die Koordinaten der oberen linken Ecken der beiden Rechtecke, werden erfasst und jeweils an vorbestimmten Stellen der RAM **16** als $t_i(X_i, Y_i)$ und $T_i(X_{T_i}, Y_{T_i})$ gespeichert.

[0063] Anschließend wird eine Differenz zwischen dem relativen Abstand Δ_i von t_i nach t_1 und dem relativen Abstand von T_i nach T_1 berechnet, basierend auf der folgenden Gleichung für alle möglichen Werte von i ($= 2, 3, 4, 5$):

$$\Delta_i = |(X_i - X_1) - (X_{T_i} - X_{T_1})|, (Y_i - Y_1) - (Y_{T_i} - Y_{T_1})|$$

[0064] Dann wird bestimmt, ob oder ob nicht alle der berechneten Werte Δ_i innerhalb eines vorbestimmten Bereichs liegen. Wenn alle Werte innerhalb des vorbestimmten Bereichs liegen, kann bestimmt werden, dass die registrierte Sprache mit der unbekannten Sprache übereinstimmt, und andernfalls wird bestimmt, dass die unbekannte Sprache nicht mit der registrierten Sprache übereinstimmt, gefolgt von einer Ausgabe des Bestimmungsergebnisses auf der Vergleichsergebnis-Ausabeeinheit **28**. Es sollte angemerkt werden, dass der hierbei benutzte vorbestimmte Bereich als Bereich definiert ist, in welchem basierend auf einer Verteilung des Ergebnisses von aus Bilddaten von Soundspekrogrammen, die von einer Vielzahl von Personen erfasst wurden, tatsächlich ("actually") berechneten Δ_i eine gewünschte Vergleichsgenauigkeit gewährleistet ist.

[0065] Auf diese Weise wird die Evaluierung gemacht bezüglich ob oder ob nicht die unbekannte Person, die die unbekannten Sprachsignaldaten $b(n)$ eingegeben hat, die Identität mit dem registrierten Sprecher hat, der die registrierten Sprachsignaldaten $a(n)$ registriert hatte, und das Evaluierungsergebnis wird ausgegeben oder gedruckt.

[0066] In der vorangehenden Sprachregistrierungs-/vergleichtsverarbeitung werden die Schablonen auf dem Soundspektrum "A" der registrierten Sprache platziert. Alternativ können die Schablonen auf dem Soundspektrum "B" der unbekannten Sprache platziert werden, um die maximal korrelierten Teilflächen T_i auf dem Soundspektrum "A" zu finden. Ferner, während die jeweiligen Teilflächen t_1 , T_1 , t_2 bis t_5 und T_2 bis T_5 als rechteckig definiert sind, sind diese Teilflächen nicht auf rechteckig beschränkt sondern können eine beliebige Gestalt haben. Ferner, während t_1 und T_1 und t_2 bis t_5 und T_2 bis T_5 bevorzugt von derselben Größe und Gestalt sind, ist ein geringfügiger Unterschied gewährbar, solange für den Sprachvergleich eine erforderliche Vergleichsgenauigkeit gewährleistet werden kann.

[0067] Weiterhin kann eine Vielzahl von anderen Verfahren als das Verfahren der Evaluierung von Δ_i in der vorangehenden Verarbeitung für die Vergleichsbestimmung verwendet werden. Zum Beispiel kann die Bestimmung gemacht werden, basierend auf einem Unterschied in Gestalt oder Fläche zwischen einer mit t_1 als Scheitelpunkt gebildeten Figur und einer mit T_1 als Scheitelpunkt gebildeten Figur.

[0068] Als Nächstes wird die in der Sprachregistrierungs-/vergleichsverarbeitung benutzte Berechnung eines Korrelationskoeffizienten beschrieben werden. Im Speziellen wird hierbei die Berechnung eines Korrelationskoeffizienten zwischen einer rechteckigen Teilfläche "A" und einer rechteckigen Teilfläche "B" beschrieben.

[0069] Als erstes wird angenommen, dass in der rechteckigen Teilfläche "A" und der rechteckigen Teilfläche "B" enthaltene Pixel jeweils $A(i, j)$ und $B(m, n)$ sind. Es sollte angemerkt werden, dass eine Gesamtzahl von in der rechteckigen Teilfläche "A" und der rechteckigen Teilfläche "B" enthaltenen Pixeln gleich ist. Ferner wird angenommen, dass Signalintensitäten, die durch Multilevelabstufungswerte, die für den Kontrast dieser Pixel kennzeichnend sind, jeweils X_{ij} und Y_{mn} sind.

[0070] Die folgende Gleichung wird definiert, wenn diese Signalintensitäten verallgemeinert und durch Z_{pq} repräsentiert werden:

$$\langle Z \rangle = N^{-1} \sum_{pq} Z_{pq}$$

[0071] In der obigen Gleichung gibt N die Gesamtzahl von in einer zugehörigen rechteckigen Teilfläche enthaltenen Pixeln an. Ferner gibt $\sum$ in der obigen Gleichung die Gesamtsumme über alle Pixel an, die in der zugehörigen rechteckigen Teilfläche enthalten sind. In anderen Worten gibt die obige Gleichung einen Mittelwert der Signalintensitäten für die in der zugehörigen rechteckigen Teilfläche enthaltenen Pixel an.

[0072] Als Nächstes wird ferner die folgende Gleichung definiert:

$$\langle Z^2 \rangle = N^{-1} \sum_{pq} Z_{pq}^2$$

[0073] Die obige Gleichung gibt einen quadratischen Mittelwert der Signalintensitäten für die in der zugehörigen rechteckigen Teilfläche enthaltenen Pixel an.

[0074] Ein Korrelationskoeffizient C_{AB} zwischen der rechteckigen Teilfläche "A" und der rechteckigen Teilfläche "B" kann hier durch die folgende Gleichung berechnet werden, welche unter Benutzung der Definitionen der vorangehenden Gleichungen ausgedrückt ist:

$$C_{AB} = \frac{\langle XY \rangle - \langle X \rangle \langle Y \rangle}{\sqrt{(\langle X^2 \rangle - \langle X \rangle^2)(\langle Y^2 \rangle - \langle Y \rangle^2)}}$$

wobei $\langle XY \rangle = (1/N) \sum_{ij} X_{ij} Y_{mn}$ ist.

[0075] Der Korrelationskoeffizient zwischen den Teilflächen wird berechnet, wobei die obige Gleichung benutzt wird.

[0076] In der Berechnung des Korrelationskoeffizienten durch die vorangehende Gleichung kann anstatt der Signalintensitäten aller Pixel in der rechteckigen Teilfläche zu benutzen die Berechnung gemacht werden, wobei nur diejenigen Pixel, welche sich auf einer beliebigen Linie innerhalb jeder der rechteckigen Teilflächen befinden; nur diejenigen Pixel, welche in einem Teil jeder der rechteckigen Teilflächen enthalten sind; oder nur diejenigen Pixel, die als Ergebnis von beliebig aus jeder der rechteckigen Teilflächen abgetasteten Pixeln ausgewählt wurden, benutzt werden, solange für den Sprachvergleich eine erforderliche Vergleichsgenauigkeit gewährleistet werden kann. Die Benutzung eines solchen Berechnungsansatzes ist vorteilhaft in einer Reduktion der Anzahl von Pixeln, die der Berechnung des Korrelationskoeffizienten unterzogen werden, und einer resultierenden Reduktion der Menge von Berechnungen. Weiterhin können alternativ andere Korrelationskoeffizienten-Berechnungsverfahren in der Sprachregistrierungs-/vergleichsverarbeitung verwendet werden.

[0077] Wie im Vorangehenden gewürdigt wird entsprechend der Sprachvergleichsvorrichtung der wie oben beschrieben konfigurierten ersten Ausgestaltung die Verarbeitung für die Konvertierung von Sprachsignalen in ein Soundspektrum nur als die Vorverarbeitung für die Berechnung einer charakteristischen Größe benötigt, die auf die Vergleichsverarbeitung angewandt wird für die tatsächliche Bestimmung der Identität zwischen registrierten Sprachsignalen und unbekannten Sprachsignalen, so dass die Anzahl von Verarbeitungsschritten signifikant reduziert werden kann. Zusätzlich, da das Soundspektrum, das aus den Sprachsignalen konvertiert wurde, als zweidimensionale Multilevelabstufungsdaten gehandhabt wird, d. h. als Pixeldaten, und ein Kontrastmuster des Soundspektrums selbst als die charakteristische Größe für die Benutzung im Vergleich benutzt wird, kann der Vergleich erreicht werden, indem eine charakteristische

Größe benutzt wird, von der personengebundene Information nicht verloren geht. Es ist daher möglich, eine Sprecheridentifizierungsvorrichtung zu realisieren, welche eine genügend hohe Trefferquote aufrecht erhält, während sie die Anzahl von Verarbeitungsschritten reduziert, um die gesamte Vergleichsverarbeitung zu vereinfachen.

[0078] In der Sprachvergleichsvorrichtung der ersten Ausgestaltung wurden Gesamtmultilevelabstufungsbilddaten eines Soundspektrogramms in der Registrierter-Sprecher-Informationsspeichereinheit **23** für eine Registrierung als eine charakteristische Größe eines registrierten Sprechers gespeichert. Für das Vergleichen unbekannter Sprachsignalen, die durch einen unbekannten Sprecher eingegeben wurden, platziert die Schablonenplatzierungseinheit **25** Schablonen im Soundspektrogramm, das aus der Registrierter-Sprecher-Informationsspeichereinheit **23** gelesen wurde, für ein Vergleichen mit Multilevelabstufungsbilddaten eines Soundspektrogramms einer unbekannten Sprache. Alternativ, wie in der folgenden Sprachvergleichsvorrichtung entsprechend einer zweiten Ausgestaltung beschrieben, können die Schablonen vorher auf den Multilevelabstufungsbilddaten eines Soundspektrogramms einer registrierten Sprache platziert werden, und nur Bilddaten, die den Teilflächen der Schablonen entsprechen, werden in der Registrierter-Sprecher-Informationsspeichereinheit **23** als eine charakteristische Größe des registrierten Sprechers gespeichert, um die Menge an Information pro registriertem Sprecher zu reduzieren, welche in der Registrierter-Sprecher-Informationsspeichereinheit **23** gespeichert werden soll.

Zweite Ausgestaltung

[0079] Fig. 7 ist ein Blockdiagramm, das die Konfiguration operationeller Funktionen in der Ausführung eines Sprachregistrierungs-/vergleichsverarbeitungsprogramms entsprechend der zweiten Ausgestaltung der Sprachvergleichsvorrichtung der vorliegenden Erfindung illustriert.

[0080] Fig. 8 ist ein Flussdiagramm, das die Sprachregistrierungs-/vergleichsverarbeitung in der zweiten Ausgestaltung der Sprachvergleichsvorrichtung illustriert.

[0081] Im Speziellen, für eine Registrierung einer charakteristischen Größe eines registrierten Sprechers, platziert die Schablonenplatzierungseinheit **25** vorher Schablonen t_i auf einem Soundspektrogramm "A" des registrierten Sprechers (Schritt A3), welches durch die Soundspektrogramm-Konvertierungseinheit **22** konvertiert wurde (Schritte A1 und A2), und jede der Teilflächen im Soundspektrogramm "A", die von jeder aller Schablonen t_i , die darauf platziert sind, umgeben ist, wird in der Registrierter-Sprecher-Informationsspeichereinheit **23** als eine charakteristische Größe des registrierten Sprechers registriert (Schritt A4).

[0082] Für das Vergleichen von Sprachdaten eines zu identifizierenden Sprechers liest die Maximal-Korrelierte-Teilflächen-Ermittlungseinheit **26** in der Vergleichseinheit **24** die Anteile im Soundspektrogramm "A" des registrierten Sprechers, die den jeweiligen Schablonen t_i entsprechen und in der Registrierter-Sprecher-Informationsspeichereinheit **23** registriert sind. Dann, ähnlich der ersten Ausgestaltung, wird der Vergleich durchgeführt, indem die jeweiligen Maximalkorrelationskoeffiziententeilflächen T_i auf dem Soundspektrogramm einer unbekannten Sprache ermittelt werden (Schritte A1 und A5 bis A8), und indem ein Lageverhältnis der jeweiligen Schablonen t_i mit einem Lageverhältnis der jeweiligen ermittelten Teilflächen T_i für die Ermittlung eines Unterschieds dazwischen verglichen wird (Schritt A9).

[0083] Wie gewürdigt, entsprechend der Sprachvergleichsvorrichtung der wie oben beschrieben konfigurierten zweiten Ausgestaltung, ist es möglich, die Menge an Information pro registriertem Sprecher, die in der Registrierter-Sprecher-Informationsspeichereinheit **23** zu speichern ist, zu reduzieren und entsprechend die Gesamtspeicherkapazität der Registrierter-Sprecher-Informationsspeichereinheit **23** selbst zu reduzieren. In anderen Worten, die Sprachvergleichsvorrichtung der zweiten Ausgestaltung kann mit einer kleineren Speicherkapazität implementiert werden als die Sprachvergleichsvorrichtung der ersten Ausgestaltung.

[0084] In der Sprachvergleichsvorrichtung entsprechend der ersten und zweiten Ausgestaltungen werden eine Vielzahl von Schablonen t_i an beliebigen Positionen in dem gesamten Soundspektrogramm "A" eines registrierten Sprechers platziert. Alternativ, wie in der folgenden Sprachvergleichsvorrichtung einer dritten Ausgestaltung beschrieben, kann die Sprachvergleichsvorrichtung so konfiguriert sein, dass sie einen Sprachabschnitt, welcher genügend Sprachcharakteristika eines registrierten Sprechers enthält, von dem Soundspektrogramm "A" einer registrierten Sprache über einen vorher gesetzten Schwellwert ermittelt und Schablonen t_i an beliebigen Positionen im im ermittelten Sprachabschnitt enthaltenen Spektrogramm "A" platziert, um eine unbekannte Sprache, basierend auf mehr charakteristischen Sprachdaten des registrierten Sprechers vergleicht, und dabei die Trefferquote weiter verbessert.

[0085] Fig. 9 ist ein Blockdiagramm, das die Konfiguration operationeller Funktionen in der Ausführung eines Sprachregistrierungs-/vergleichsverarbeitungsprogramms entsprechend einer dritten Ausgestaltung der Sprachvergleichsvorrichtung der vorliegenden Erfindung illustriert.

[0086] Die operationellen Funktionsblöcke der Sprachvergleichsvorrichtung in der dritten Ausgestaltung enthalten zusätzlich eine Sprachabschnittsermittlungseinheit **29**, die zwischen der Soundspektrum-Konvertierungseinheit **22** und der Schablonenplatzierungseinheit **25** bereitgestellt wird.

[0087] Die Sprachabschnittsermittlungseinheit **29** ermittelt einen Sprachabschnitt, welcher in ausreichendem Maße Sprachcharakteristika eines registrierten Sprechers enthält, aus einem Soundspektrum "A" einer registrierten Sprache, das durch die Soundspektrum-Konvertierungseinheit **22** konvertiert wurde (siehe Fig. 11). Im Sprachabschnitt auf dem Soundspektrum "A", der durch die Sprachabschnittsermittlungseinheit **29** ermittelt wurde, platziert die Schablonenplatzierungseinheit **25** eine Vielzahl von Schablonen t_i , welche als Basis für den Sprachvergleich dienen.

[0088] Fig. 10 ist ein Flussdiagramm, das eine Sprachregistrierungs-/vergleichsverarbeitung in der dritten Ausgestaltung der Sprachvergleichsvorrichtung illustriert.

[0089] Fig. 11 zeigt, wie ein Sprachabschnitt aus dem Soundspektrum "A" der registrierten Sprache in der Sprachabschnittsermittlungseinheit **29** in der Sprachvergleichsvorrichtung der dritten Ausgestaltung ermittelt wird.

[0090] Fig. 12A und 12B zeigen einen Vergleich eines registrierten Sprachspektrums mit einem (zu vergleichenden) unbekannten Sprachspektrum und einer Platzierung einer Schablone auf dem registrierten Sprachspektrum entsprechend der Sprachregistrierungs-/vergleichsverarbeitung in der Sprachvergleichsvorrichtung der dritten Ausgestaltung.

[0091] Fig. 13A und 13B zeigen einen Vergleich einer Verteilung von Positionen, an welchen Schablonenteilflächen auf einem registrierten Sprachspektrum platziert sind, mit einer Verteilung von Positionen, an welchem maximal korrelierte Teilflächen auf einem (zu vergleichenden) unbekannten Sprachspektrum ermittelt werden, entsprechend der Sprachregistrierungs-/vergleichsverarbeitung in der Sprachvergleichsvorrichtung der dritten Ausgestaltung.

[0092] Für die Benutzung der Sprachvergleichsvorrichtung als Personenidentifikationsvorrichtung werden Sprachsignaldaten einer registrierten Person, die der Identifikation unterzogen wird, zuerst von der Sprachsignal-Eingabeeinheit **21** eingegeben. Diese eingegebenen registrierten Sprachsignaldaten $a(n)$ werden in der Soundspektrum-Konvertierungseinheit **22** in ein Soundspektrum "A" konvertiert (Schritte B1 und B2).

[0093] Dann ermittelt die Sprachabschnittsermittlungseinheit **29** einen Sprachabschnitt aus dem Soundspektrum "A" der registrierten Sprache, die durch die Soundspektrum-Konvertierungseinheit **22** konvertiert wurde (Schritt B3).

[0094] Hier, wie in Fig. 11 illustriert, summiert die Sprachabschnittsermittlungsverarbeitung, die in der Sprachabschnittsermittlungseinheit **29** durchgeführt wird, zuerst für das Soundspektrum "A" der registrierten Sprache Spektralwerte in der Richtung der Frequenzachse auf, um eine additive Spektralverteilung zu berechnen. Als Nächstes wird ein Abschnitt innerhalb der additiven Spektralverteilung, der alle Bereiche enthält, die größere Werte als ein zuvor gesetzter Schwellwert haben, als Sprachabschnitt ermittelt. Indes, der gesetzte Schwellwert ist als Hälfte eines Maximalwerts in der additiven Spektralverteilung definiert.

[0095] Nachdem der Sprachabschnitt auf diese Weise in dem Soundspektrum "A" ermittelt wurde, definiert die Schablonenplatzierungseinheit **25** eine Vielzahl von rechteckigen Teilflächen, d. h. Schablonen t_i ($i = 1$ bis m , wobei m eine ganze Zahl größer oder gleich zwei ist) und platziert die Schablonen t_i auf dem Sprachabschnitt im Soundspektrum "A", wie in Fig. 12A gezeigt (Schritt B4).

[0096] Dann wird eine partielle Teilfläche A_t , die alle Schablonen t_i enthält, die auf dem Sprachabschnitt im Soundspektrum "A" der registrierten Sprache platziert wurden, in der Registrierter-Sprecher-Informationsspeichereinheit **23** für die Registrierung als eine charakteristische Größe des registrierten Sprechers ge-

speichert (Schritt B5).

[0097] Andererseits, nachdem Sprachsignalen eines zu identifizierenden Sprechers von der Sprachsignalen-Eingabeeinheit **26** für die Identifikation in der Personenidentifizierungsvorrichtung eingegeben wurden, werden die eingegebenen unbekannten Sprachsignalen $b(n)$ in der Soundspektrogramm-Konvertierungseinheit **22** wie oben beschrieben in ein Soundspektrogramm "B" konvertiert (Schritte B1 bis B6).

[0098] Anschließend wird die partielle Teilfläche A_t , die die jeweiligen Schablonen t_i im Soundspektrogramm "A", das in der Registrierter-Sprecher-Informationsspeichereinheit **23** als die charakteristische Größe des registrierten Sprechers registriert wurde, enthält, in die Maximal-Korrelierte-Teilflächen-Ermittlungseinheit **26** in der Vergleichseinheit **24** eingelesen. Die Maximal-Korrelierte-Teilflächen-Ermittlungseinheit **26** rasterscannt Bilddaten einer Teilfläche, die jeder der Schablonen t_i entspricht, auf einer Pixel für Pixel-Basis auf dem Soundspektrogramm "B" der unbekannten Sprache, um sequenziell einen Korrelationskoeffizienten zu berechnen, indem sie alle Pixeldaten in der zugehörigen Schablone t_i und Pixeldaten im Soundspektrogramm "B" der entsprechenden unbekannten Sprache benutzt (Schritt B7). Dann wird eine Teilfläche T_i im Soundspektrogramm "B" einer unbekannten Sprache ermittelt, in welcher ein maximaler Korrelationskoeffizient vorhanden ist, wie in **Fig. 13A** und **13B** gezeigt (Schritt B8).

[0099] Die Berechnung des maximalen Korrelationskoeffizienten auf dem Soundspektrogramm "B" der unbekannten Sprache basierend auf den Bilddaten jeder Schablone t_i , die auf dem Sprachabschnitt im Soundspektrogramm "A" der registrierten Sprache definiert ist, und die Ermittlung der Teilfläche T_i werden sequenziell für jede Schablone t_i durchgeführt (Schritte S7 bis S9). Dann, nachdem bestimmt wurde, dass die jeweiligen Teilflächen T_i auf dem Soundspektrogramm "B" der unbekannten Sprache für alle Schablonen t_i als diejenigen ermittelt worden sind, die maximale Korrelationskoeffizienten haben, vergleicht die Vergleichsbestimmungseinheit **27** in der Vergleichseinheit **24** eine Verteilung der jeweiligen Schablonenteilflächen $\{t_i\}$, die auf dem Sprachabschnitt im Soundspektrogramm "A" der registrierten Sprache platziert sind (Lagebeziehung), mit einer Verteilung der jeweiligen Teilflächen $\{T_i\}$, die von dem Soundspektrogramm "B" der unbekannten Sprache basierend auf ihren jeweiligen Koordinatendaten ermittelt wurden (Lagebeziehung), um die Identität dazwischen zu evaluieren (Schritte B9 und B10).

[0100] Auf diese Weise wird die Evaluierung gemacht, ob oder ob nicht die unbekannte Person, die die unbekannten Sprachsignalen $b(n)$ eingegeben hat, die Identität mit dem registrierten Sprecher hat, der die registrierten Sprachsignalen $a(n)$ registriert hatte, und das Evaluationsergebnis wird ausgegeben oder gedruckt.

[0101] Wie gewürdigt, entsprechend der Sprachvergleichsvorrichtung der wie oben beschrieben konfigurierten dritten Ausgestaltung, werden nur zwei Verarbeitungsschritte, d. h. die Verarbeitung für die Konvertierung eines Soundspektrogramms aus Sprachsignalen und die Sprachabschnittsermittlungsverarbeitung für die Ermittlung eines Sprachabschnitts aus dem Soundspektrogramm einer registrierten Sprache, als die Vorverarbeitung für die Berechnung der charakteristischen Größe benötigt, die auf die Vergleichsverarbeitung angewandt wird für die tatsächliche Bestimmung der Identität zwischen registrierten Sprachsignalen und unbekannten Sprachsignalen, wobei sie es möglich machen, die Anzahl von Verarbeitungsschritten zu reduzieren. Zusätzlich wird das aus den Sprachsignalen konvertierte Soundspektrogramm als zweidimensionale Multilevelabstufungsdaten gehandhabt, d. h. als Pixeldaten, und ein Kontrastmuster des Soundspektrogramms selbst wird als die charakteristische Größe für die Benutzung in dem Vergleich benutzt, und die Schablonen werden definiert und auf einem Sprachabschnitt, der durch die Sprachabschnittsermittlungsverarbeitung ermittelt wird, platziert, um ein Kontrastmuster in den Schablonen für den Vergleich als eine charakteristische Größe der registrierten Sprache zu benutzen. Deshalb kann der Vergleich erreicht werden, indem eine charakteristische Größe benutzt wird, welche personengebundene Information vollständiger widerspiegelt als die Sprachvergleichsvorrichtung, die in der ersten und zweiten Ausgestaltung beschrieben wurde. Es ist daher möglich, eine Sprecheridentifizierungsvorrichtung zu realisieren, welche die Trefferquote gegenüber den ersten und zweiten Ausgestaltungen weiter verbessert, während sie die Anzahl von Verarbeitungsschritten reduziert, um die Gesamtvergleichsverarbeitung zu vereinfachen.

[0102] Zusätzliche Vorteile und Modifikationen werden Fachmännern leicht einfallen. Daher ist die vorliegende Erfindung in ihren breiteren Aspekten nicht auf die speziellen Details, repräsentativen Geräte und hierin gezeigten und beschriebenen illustrierten Beispiele limitiert. Entsprechend können mannigfaltige Modifikationen gemacht werden, ohne den Schutzbereich des generellen erfinderischen Konzepts, wie in den beigefügten Ansprüchen definiert, zu verlassen. Zum Beispiel können die in den jeweiligen Ausgestaltungen beschriebenen Techniken, d. h. die Sprachregistrierungs-/vergleichsverarbeitung in der ersten Ausgestaltung, die im Flusssdi-

agramm der **Fig. 3** illustriert ist; die Sprachregistrierungs-/vergleichtsverarbeitung in der zweiten Ausgestaltung, die im Flussdiagramm der **Fig. 8** illustriert ist; die Sprachregistrierungs-/vergleichtsverarbeitung in der dritten Ausgestaltung, die im Flussdiagramm der **Fig. 10** illustriert ist, und Ähnliches in einem externen Aufnahmemedium **14** gespeichert werden, wie einer Speicherkarte (ROM-Karte, RAM-Karte oder Ähnliches), einer magnetischen Platte (Diskette, Festplatte oder Ähnliches), einer optischen Platte (CD-ROM, DVD oder Ähnliches), einem Halbleiterspeicher oder Ähnlichem für die Verbreitung als ein Programm, das einen Computer veranlassen kann, diese Techniken auszuführen. In diesem Fall kann ein Computer, der als die Sprachvergleichsvorrichtung funktioniert, das auf dem externen Aufzeichnungsmedium **14** gespeicherte Programm in das Speicherelement **13** einlesen, und die Sprachregistrierungs- und -vergleichsfunktionen implementieren, die in den jeweiligen Ausgestaltungen beschrieben wurden und deren Betrieb durch das Leseprogramm kontrolliert wird, um eine Verarbeitung auszuführen, die denen ähnlich ist, die durch die zuvor genannten Techniken bereitgestellt wird.

[0103] Ferner können Daten, die durch das Programm benötigt werden, um die jeweiligen Techniken zu implementieren, über ein Netzwerk (öffentliches Netzwerk) übermittelt werden in Form von Programmcodes, so dass die Programmdateien durch den Übermittlungscontroller **15** der Sprachvergleichsvorrichtung abgerufen werden, der mit dem Netzwerk verbunden ist, um die zuvor genannte Sprachregistrierungs- und -vergleichsfunktion zu implementieren.

Gewerbliche Anwendbarkeit

[0104] Entsprechend der vorliegenden Erfindung werden eine Sprachvergleichsvorrichtung und ein Sprachvergleichsverfahren bereitgestellt, welche fähig sind, einen Sprecher mit einer hohen Trefferquote zu identifizieren, ohne das Erfordernis nach einer großen Anzahl von Vorverarbeitungsschritten.

Patentansprüche

1. Eine Sprachvergleichsvorrichtung für die Identifizierung einer Person mit Sprachdaten, umfassend: Datenkonvertierungsmittel (**22**) zum Konvertieren von zwei Sprachsignalen, die einem Vergleich unterzogen werden, in zwei zweidimensionale Daten, die kennzeichnend für Sprachcharakteristika der zwei Sprachsignale sind;

Schablonenplatzierungsmittel (**25**) zum Platzieren einer Vielzahl von Schablonen (t_1 bis t_5) zum Definieren einer Vielzahl von Teilflächen auf den einen der zweidimensionalen Daten;

Korrelierte-Teilflächen-Ermittlungsmittel (**26**) zum Ermitteln solcher Teilflächen (T_1 bis T_5) auf den anderen der zweidimensionalen Daten, die eine maximale Korrelation bezüglich einer Vielzahl von Teilflächen auf den anderen der zweidimensionalen Daten haben und der Vielzahl von Schablonen entsprechen;

Vergleichsbestimmungsmittel (**27**) zum Vergleichen einer gegenseitigen Lagebeziehung der Vielzahl von Schablonen auf den einen der zweidimensionalen Daten mit einer gegenseitigen Lagebeziehung der Teilflächen auf den anderen zweidimensionalen Daten, welche durch das Korrelierte-Teilflächen-Ermittlungsmittel ermittelt wurden, um Identität zwischen den zwei Sprachsignalen zu bestimmen.

2. Die Sprachvergleichsvorrichtung entsprechend Anspruch 1, weiterhin umfassend Registrierter-Sprecher-Informationsspeichermittel (**23**) zum Speichern zweidimensionaler Daten, die einem Sprachsignal eines registrierten Sprechers entsprechen, und wobei

das Schablonenplatzierungsmittel Mittel zum Platzieren einer Vielzahl von Schablonen auf den zweidimensionalen Daten des registrierten Sprechers, die von dem Registrierter-Sprecher-Informationsspeichermittel ausgelesen wurden, enthält; und

das Korrelierte-Teilflächen-Ermittlungsmittel Mittel enthält zum Ermitteln einer Vielzahl von Teilflächen auf zweidimensionalen Daten eines unbekannten Sprechers, die eine maximale Korrelation bezüglich einer Vielzahl von Teilflächen auf den zweidimensionalen Daten des unbekannten Sprechers haben, und der Vielzahl von Schablonen entsprechen.

3. Die Sprachvergleichsvorrichtung entsprechend Anspruch 2, wobei das Registrierter-Sprecher-Informationsspeichermittel Mittel enthält zum Speichern von Daten bezüglich der Teilflächen, die der Vielzahl von Schablonen entsprechen, die auf den zweidimensionalen Daten der registrierten Sprache durch das Schablonenplatzierungsmittel platziert wurden; und

das Korrelierte-Teilflächen-Ermittlungsmittel Mittel enthält zum Ermitteln einer Vielzahl von Teilflächen auf den zweidimensionalen Daten des unbekannten Sprechers, die eine maximale Korrelation bezüglich Teilflächen auf den zweidimensionalen Daten des unbekannten Sprechers haben, und der Vielzahl von Schablonen entsprechen, die in dem Registrierter-Sprecher-Informationsspeichermittel gespeichert sind.

4. Die Sprachvergleichsvorrichtung entsprechend einem der Ansprüche 1 bis 3, weiterhin umfassend ein Registrierter-Sprecher-Informationsspeichermittel (**23**) zum Speichern von zweidimensionalen Daten, die einem Sprachsignal eines registrierten Sprechers entsprechen, und wobei das Schablonenplatzierungsmittel Mittel enthält zum Platzieren einer Vielzahl von Schablonen auf zweidimensionalen Daten eines unbekannten Sprechers; und das Korrelierte Teilflächen-Ermittlungsmittel Mittel enthält zum Ermitteln einer Vielzahl von Teilflächen auf den zweidimensionalen Daten des unbekannten Sprechers, die eine maximale Korrelation bezüglich einer Vielzahl von Teilflächen auf den zweidimensionalen Daten des unbekannten Sprechers haben, und der Vielzahl von Schablonen entsprechen.

5. Die Sprachvergleichsvorrichtung entsprechend einem der Ansprüche 1 bis 4, weiterhin umfassend ein Sprachabschnittsermittlungsmittel (**29**) zum Ermitteln eines Sprachabschnitts in den zweidimensionalen Daten, und wobei das Schablonenplatzierungsmittel Mittel enthält zum Platzieren einer Vielzahl von Schablonen auf zweidimensionalen Daten in dem Sprachabschnitt, der durch das Sprachabschnittsermittlungsmittel ermittelt wurde.

6. Die Sprachvergleichsvorrichtung entsprechend einem der Ansprüche 1 bis 5, wobei das Datenkonvertierungsmittel das Sprachsignal in ein Soundspektrogramm konvertiert.

7. Ein Sprachdatenvergleichsverfahren zum Identifizieren einer Person mit Sprachdaten, umfassen die folgenden Schritte:

Konvertieren (S2, S4, A2, A5, B2, B6) zweier Sprachsignale, die einem Vergleich unterzogen werden, in zwei zweidimensionale Daten, die für Sprachcharakteristika der zwei Sprachsignale kennzeichnend sind;

Platzieren (S5, A3, B4) einer Vielzahl von Schablonen zum Definieren einer Vielzahl von Teilflächen auf den einen der zweidimensionalen Daten;

Ermitteln (S7, A7, B8) von Teilflächen auf den anderen der zweidimensionalen Daten, die eine maximale Korrelation bezüglich einer Vielzahl von Teilflächen auf den anderen der zweidimensionalen Daten haben und der Vielzahl von Schablonen entsprechen; und

Vergleichen (S9, A9, B10) eines gegenseitigen Lageverhältnisses der Vielzahl von Schablonen auf den einen der zweidimensionalen Daten mit einem gegenseitigen Lageverhältnis der Vielzahl von Teilflächen auf den anderen der zweidimensionalen Daten, die durch das Korrelierte-Teilflächen-Ermittlungsmittel ermittelt wurden, um Identität zwischen den beiden Sprachsignalen zu bestimmen.

8. Das Sprachvergleichsverfahren des Anspruchs 7, adaptiert zum Betreiben des Systems eines der Ansprüche 1 bis 6.

9. Ein Programmaufzeichnungsmedium, das computerlesbaren Programmcode hat, der konfiguriert ist, um jeden der Schritte des Verfahrens der Ansprüche 7 oder 8 auszuführen.

Es folgen 12 Blatt Zeichnungen

FIG.1

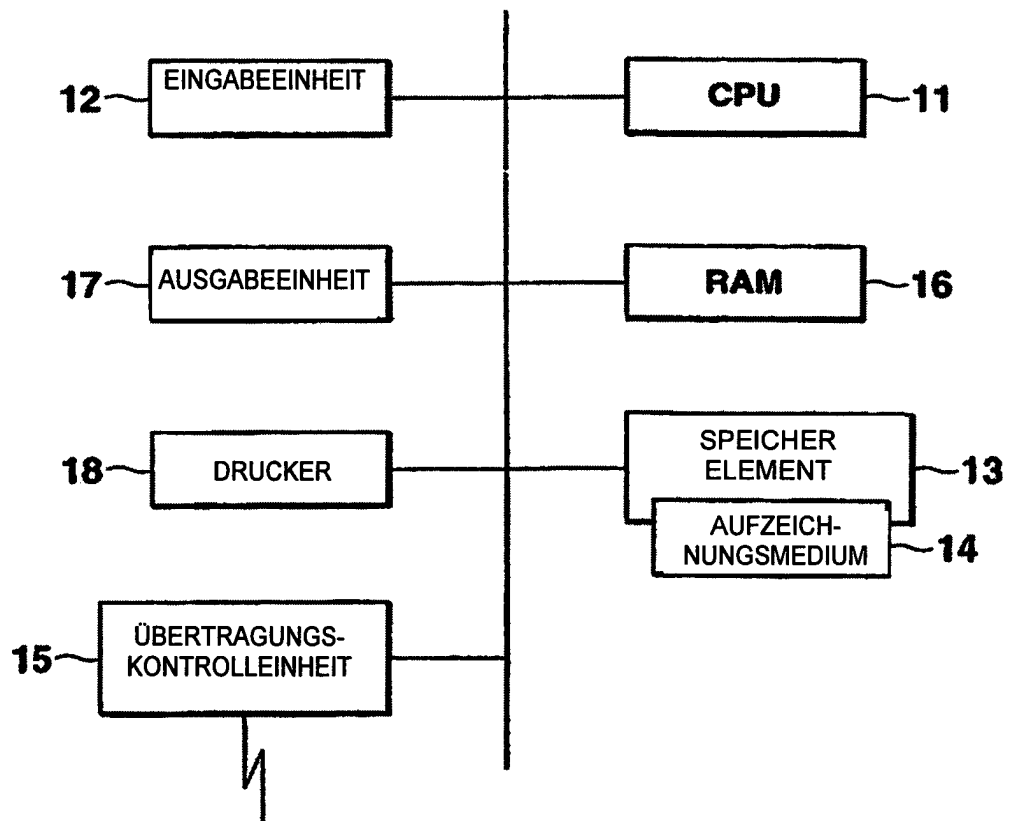


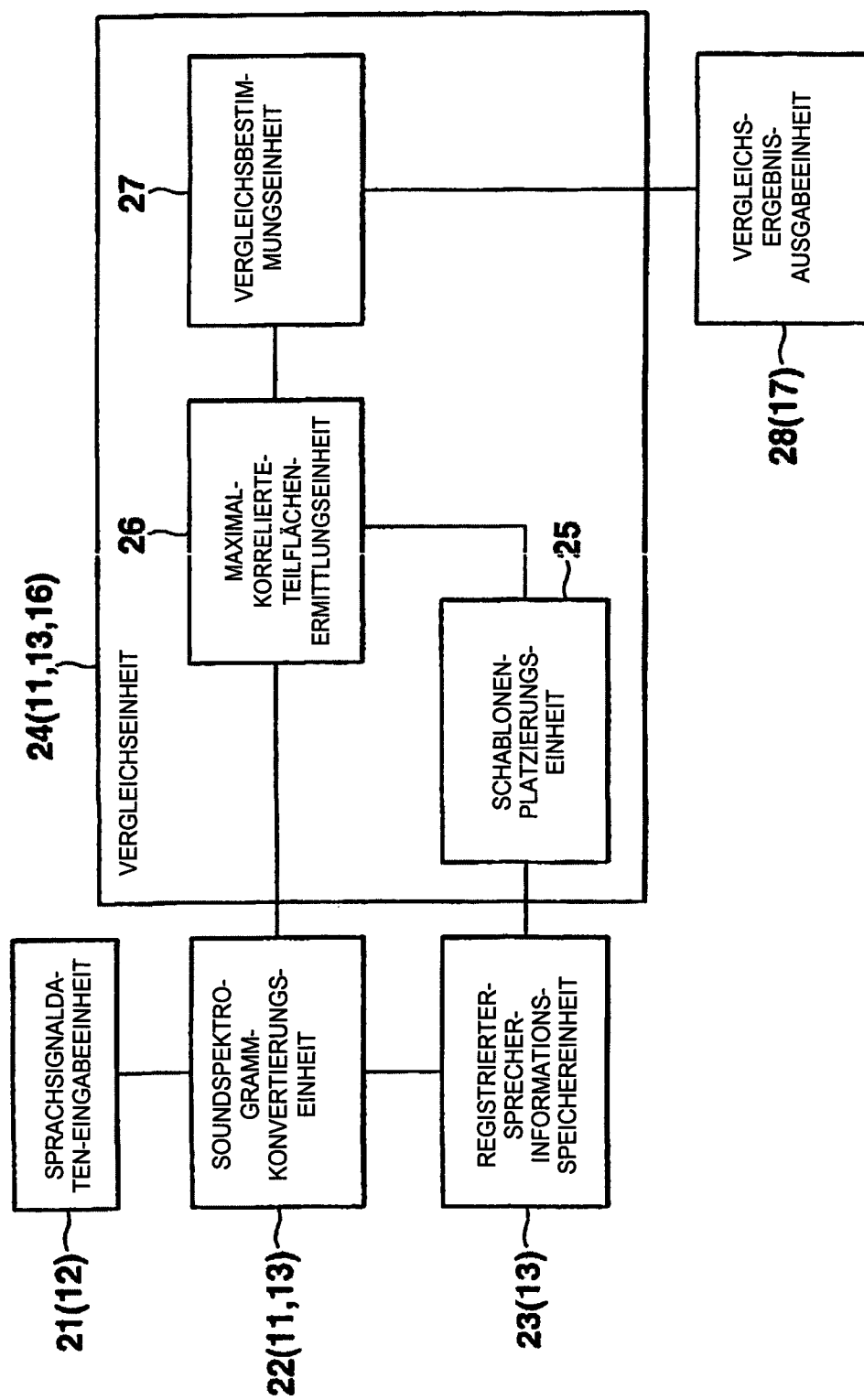
FIG.2

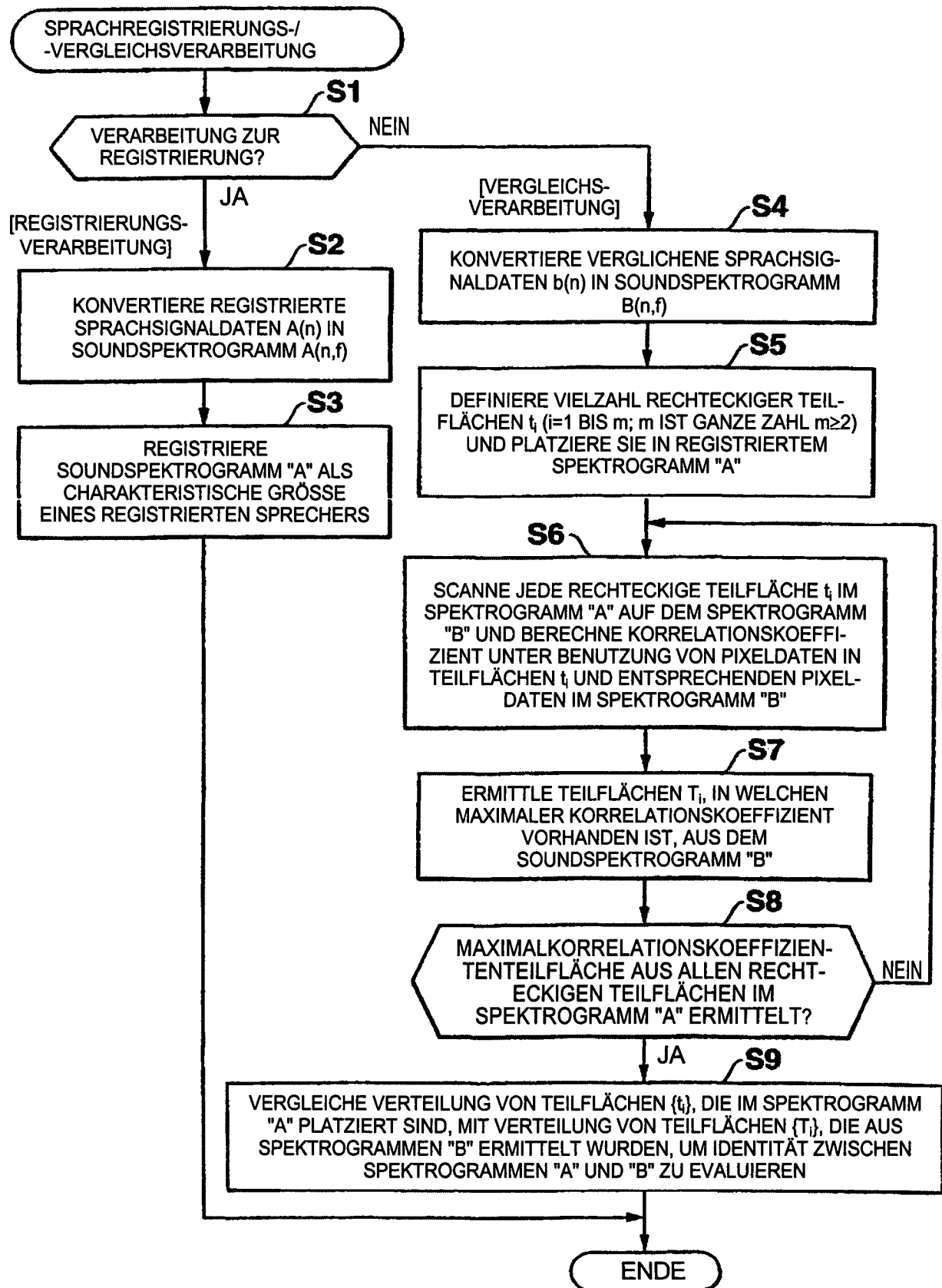
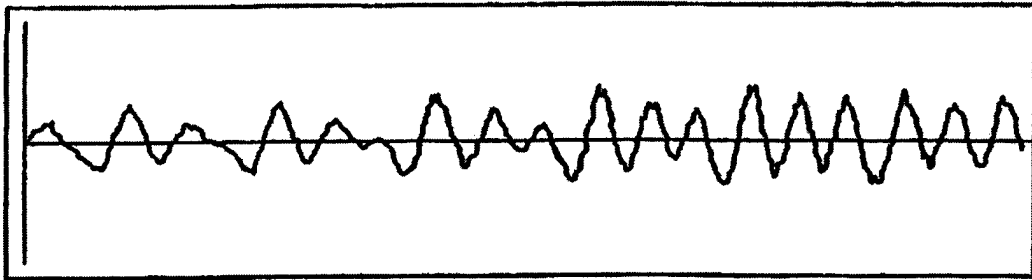
FIG.3

FIG.4A

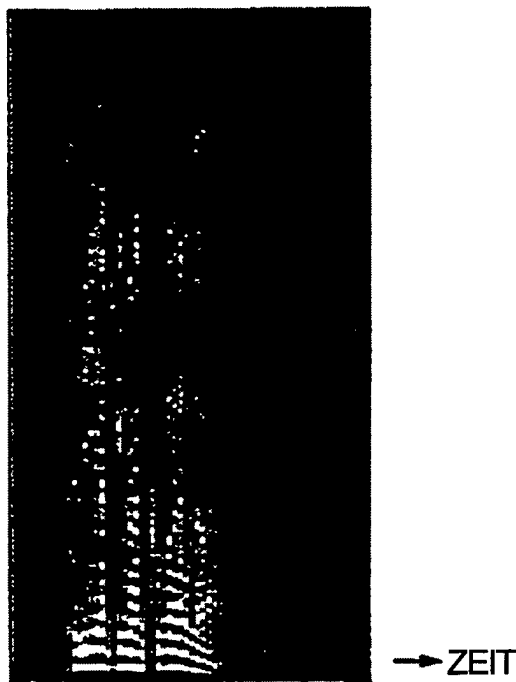


SPRACHSIGNALDATENWELLENFORM $a(n)$



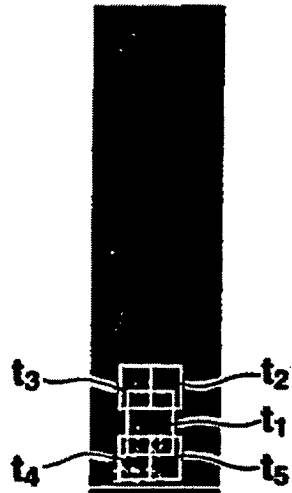
FREQUENZ ↑

FIG.4B



FREQUENZSPEKTRENZEIT-
SERIENVERTEILUNG $A(n,f)$
(SOUNDSPEKTROGRAMM)

FIG.5A



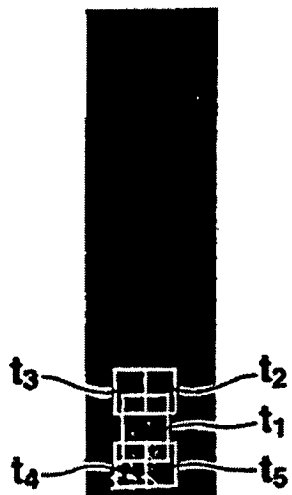
SOUNDSPEKTROGRAMM: A
(REGISTRIERTE DATEN)

FIG.5B



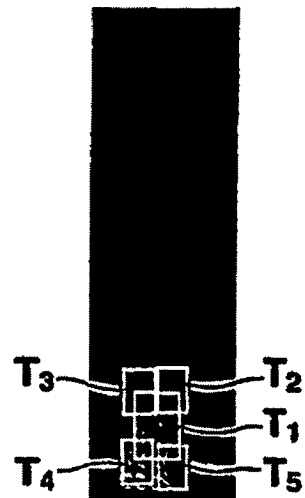
SOUNDSPEKTROGRAMM: B
(VERGLICHENE DATEN)

FIG.6A



SOUNDSPEKTROGRAMM: A
(REGISTRIERTE DATEN)

FIG.6B



SOUNDSPEKTROGRAMM: B
(VERGLICHENE DATEN)

FIG.7

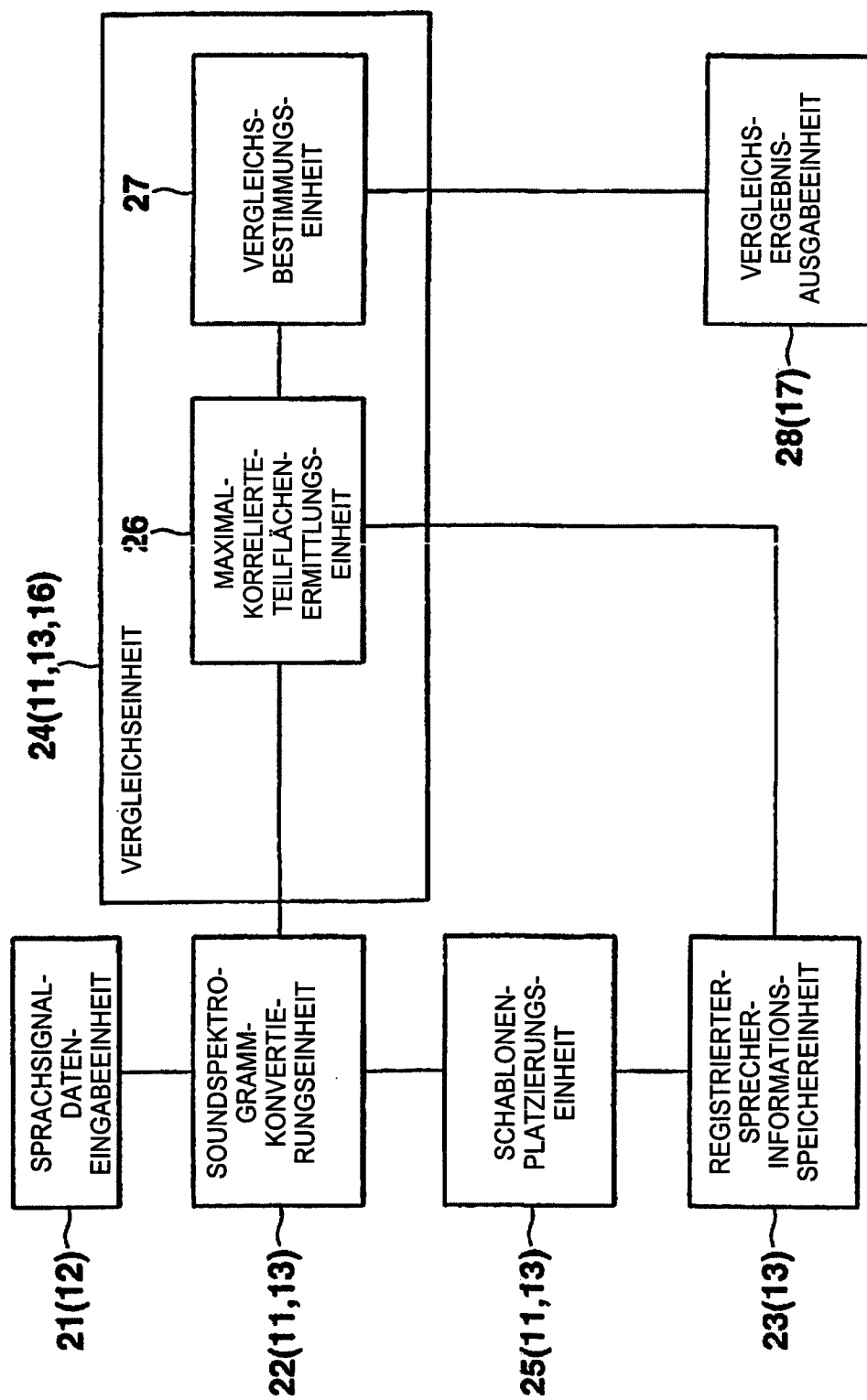


FIG.8

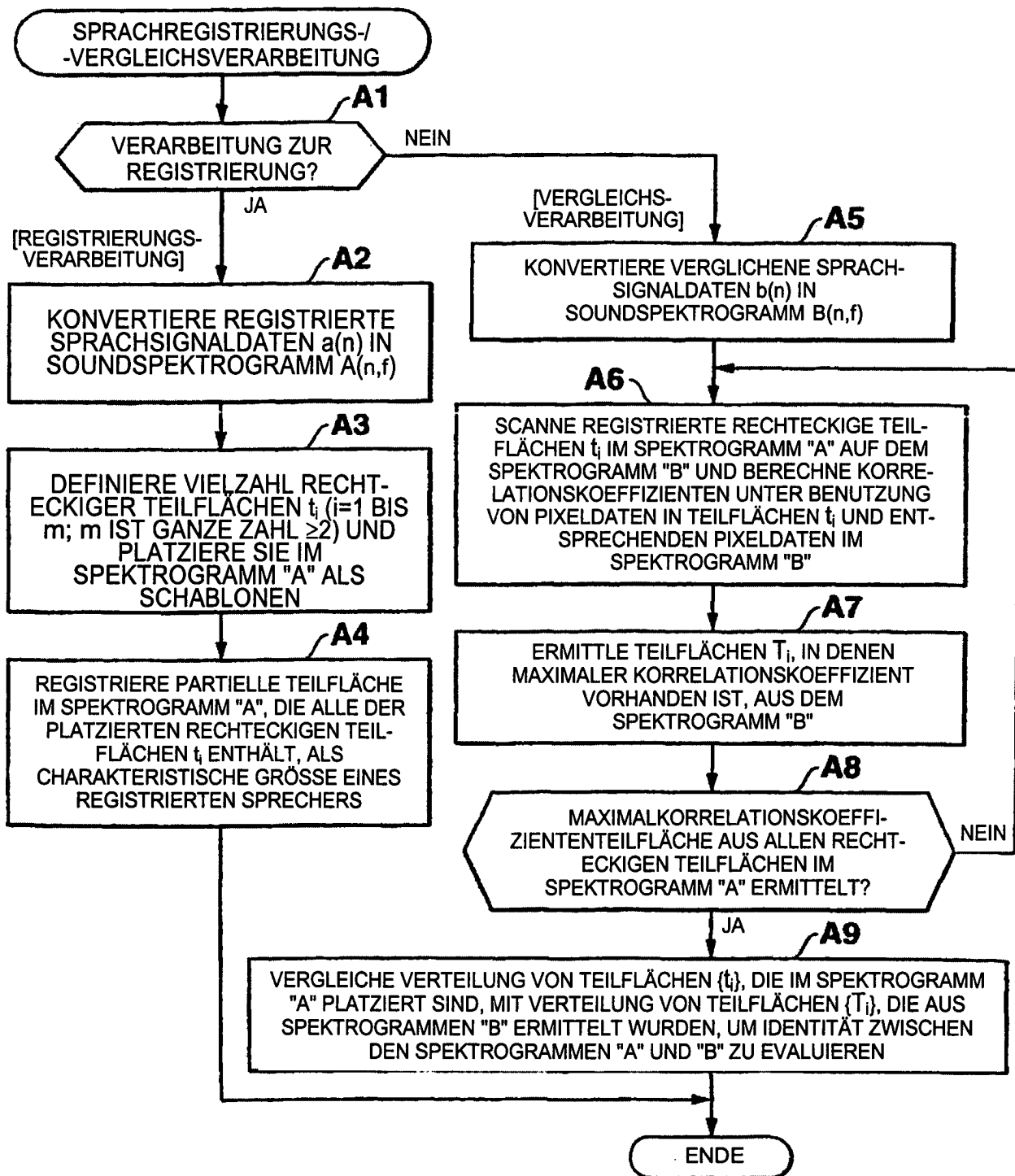


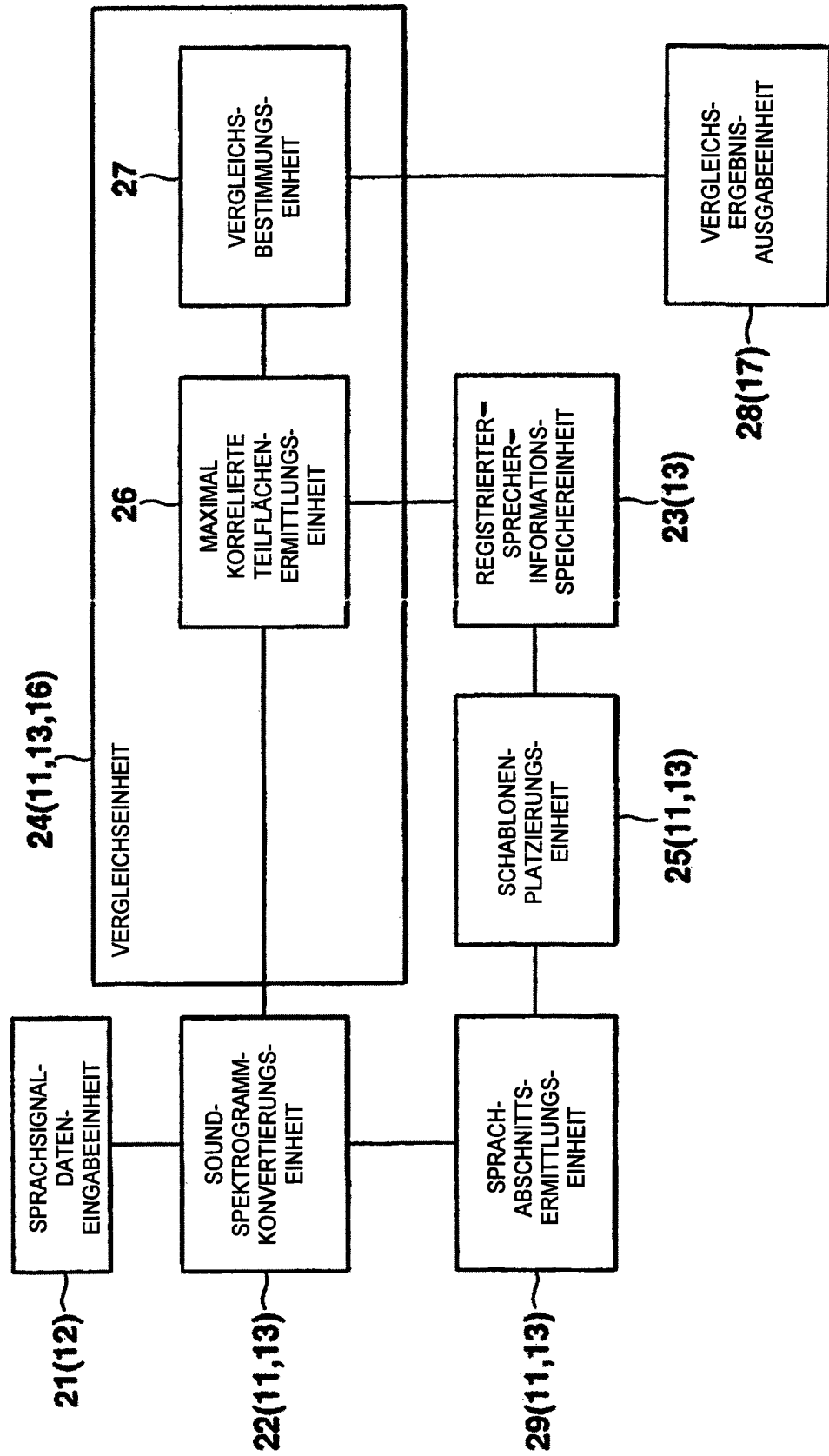
FIG.9

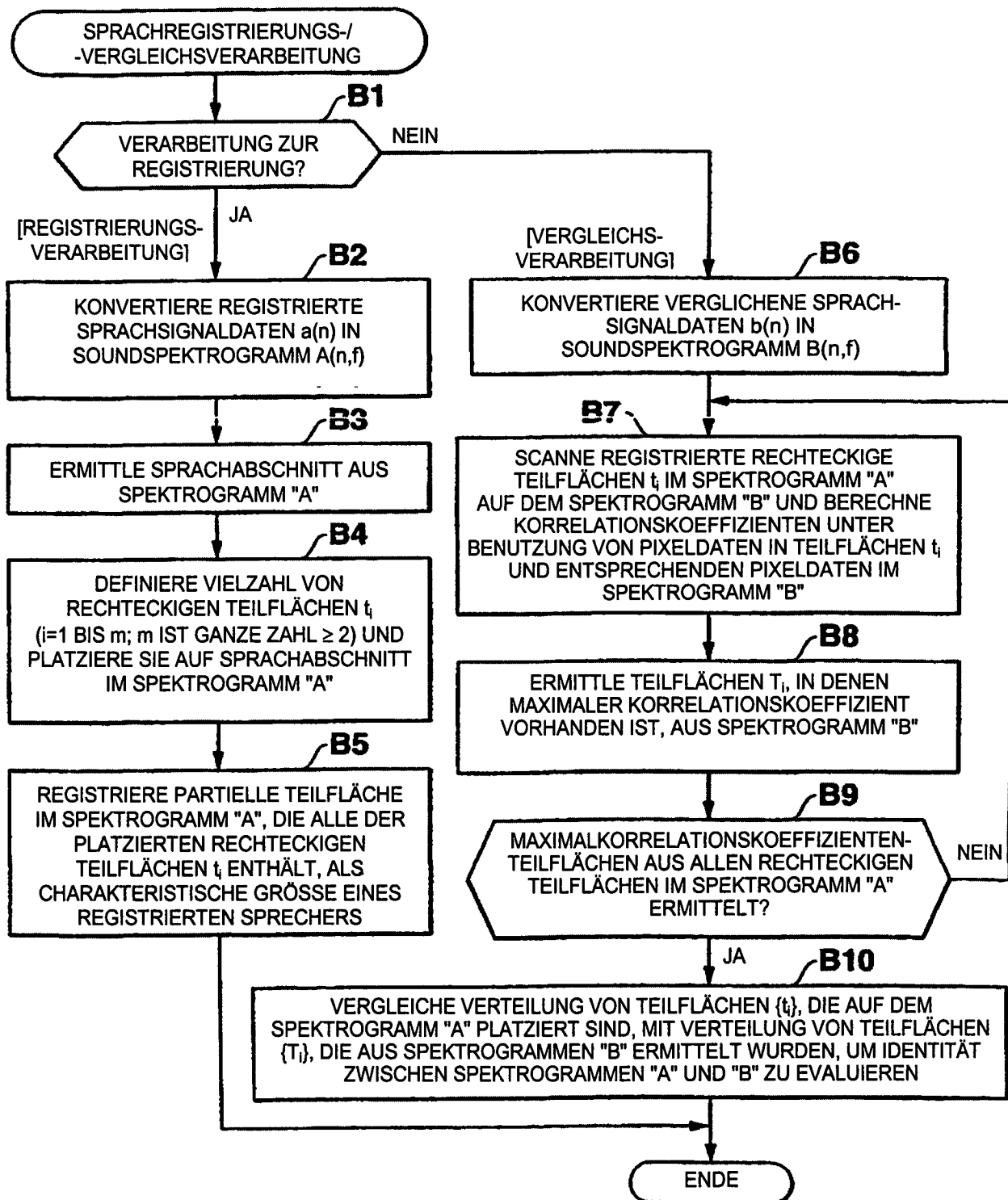
FIG.10

FIG.11

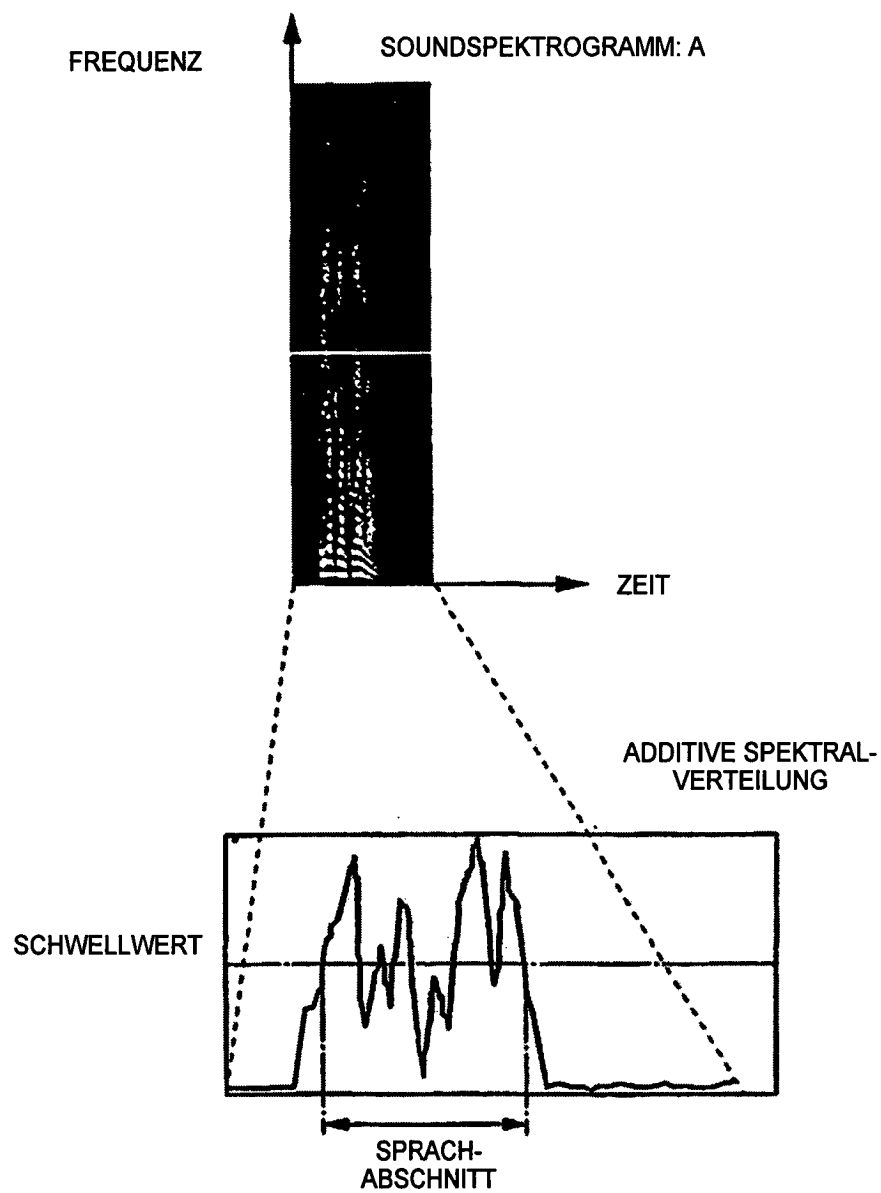
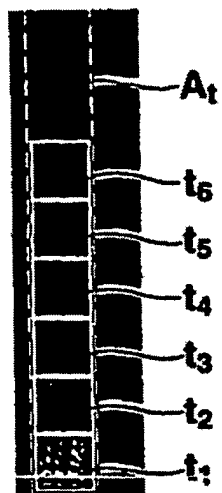


FIG.12A



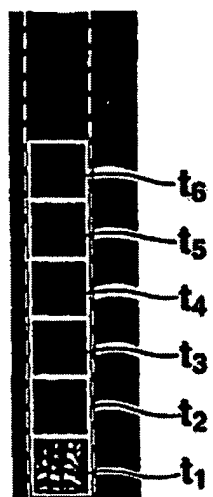
SOUNDSPEKTROGRAMM: A
(REGISTRIERTE DATEN)

FIG.12B



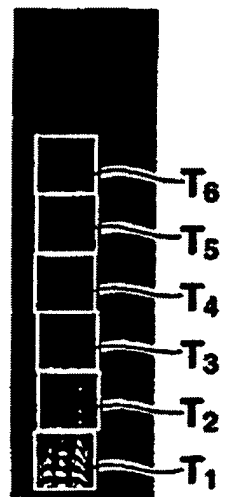
SOUNDSPEKTROGRAMM: B
(VERGLICHENE DATEN)

FIG.13A



SOUNDSPEKTROGRAMM: A
(REGISTRIERTE DATEN)

FIG.13B



SOUNDSPEKTROGRAMM: B
(VERGLICHENE DATEN)

FIG.14
STAND DER TECHNIK

