



(19)  
**Bundesrepublik Deutschland**  
**Deutsches Patent- und Markenamt**

(10) **DE 10 2006 002 603 A1 2007.07.19**

(12)

## Offenlegungsschrift

(21) Aktenzeichen: **10 2006 002 603.9**

(22) Anmeldetag: **13.01.2006**

(43) Offenlegungstag: **19.07.2007**

(51) Int Cl.<sup>8</sup>: **H04M 3/56 (2006.01)**

**G10L 17/00 (2006.01)**

**G10L 15/20 (2006.01)**

(71) Anmelder:

**Deutsche Telekom AG, 53113 Bonn, DE**

(72) Erfinder:

**Trinkel, Marian, 52372 Kreuzau, DE; Müller, Christel, 15732 Schulzendorf, DE**

(56) Für die Beurteilung der Patentfähigkeit in Betracht zu ziehende Druckschriften:

**DE 102 61 478 A1**

**DE 102 38 287 A1**

**DE 696 04 268 T2**

**GB 23 11 684 A**

**US2003/01 25 954 A1**

**US 56 31 904 A**

**EP 14 53 287 A1**

**EP 12 92 090 A1**

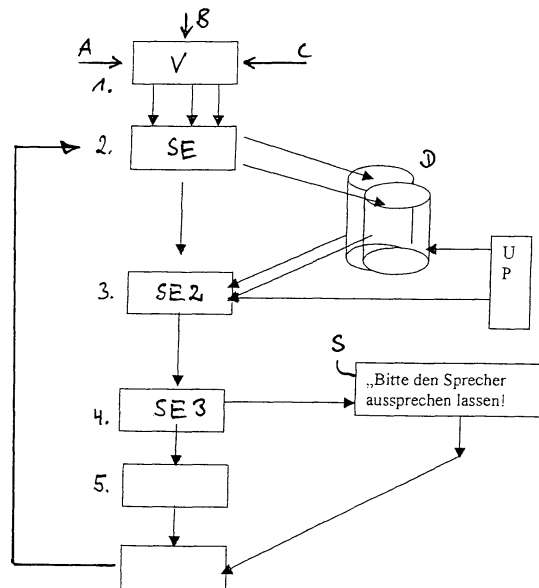
**EP 09 73 312 A1**

**Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen**

Rechercheantrag gemäß § 43 Abs. 1 Satz 1 PatG ist gestellt.

(54) Bezeichnung: **Verfahren und System zur Durchführung einer Sprachkonferenz**

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Durchführung einer Sprachkonferenz mit mehreren Teilnehmern, die miteinander kommunizieren, insbesondere über ein Telekommunikationsnetzwerk mittels Telekommunikationsgeräten, wobei die von einem Teilnehmer gesprochenen Worte mittels einer Spracherkennung erkannt und gespeichert werden und wobei bei einem zeitgleichen Sprechen von wenigstens zwei Teilnehmern eine Aufforderung zum seriellen Sprechen signalisiert wird. Die Erfindung betrifft weiterhin ein System zur Durchführung einer Sprachkonferenz mit mehreren Teilnehmern, die miteinander kommunizieren.



**Beschreibung**

**[0001]** Die Erfindung betrifft ein Verfahren und ein System zur Durchführung einer Sprachkonferenz mit mehreren Teilnehmern, die miteinander kommunizieren, insbesondere über ein Telekommunikationsnetzwerk mittels Telekommunikationsgeräten, wobei die von einem Teilnehmer gesprochenen Worte mittels einer Spracherkennung erkannt und gespeichert werden.

**[0002]** Ein derartiges Verfahren und System ist beispielsweise aus der deutschen Patentanmeldung DE 102 61 478 bekannt. Es ist beispielsweise offenbart, Sprachdaten, die während einer Konferenzschaltung entstehen, für einen späteren Abruf zu speichern.

**[0003]** Bei Konferenzen, bei denen mehrere Teilnehmer miteinander kommunizieren, ist das Problem bekannt, dass oftmals mehrere Teilnehmer gleichzeitig sprechen, sich also beispielsweise ins Wort fallen, ohne dass einer der Teilnehmer zunächst zu Ende sprechen konnte, was insbesondere eine automatische Spracherkennung zur Erkennung und Speicherung der gesprochenen Worte an ihre Grenzen bringt. So ist bislang ein zuverlässiges Erkennen gesprochener Worte nur möglich, wenn die Sprechende Person sauber und deutlich und bevorzugt mit zumindest leichten Pausen zwischen den einzelnen Worten spricht. Überlappen die Worte, insbesondere wenn mehrere Personen gleichzeitig sprechen, so kann eine Erkennung und Speicherung der Worte nicht mehr erfolgen.

**[0004]** Aufgabe der Erfindung ist es, ein Verfahren und ein System bereitzustellen, mittels dem eine optimale Spracherkennung und Speicherung erreicht wird.

**[0005]** Diese Aufgabe wird mit einem Verfahren und einem System der eingangs genannten Art gelöst, bei dem bei einem zeitgleichen Sprechen von wenigstens zwei Teilnehmern, eine Aufforderung zum seriellen Sprechen signalisiert wird.

**[0006]** Der Einsatz eines derartigen Verfahrens oder Systems ist hierbei nicht zwingend auf die Durchführung einer Sprachkonferenz beschränkt, bei der die einzelnen Teilnehmer über Telekommunikationsgeräte und somit über ein Telekommunikationsnetzwerk zusammengeschaltet sind, sondern grundsätzlich auch bei einer Sprachkonferenz einsetzbar, bei der sämtliche teilnehmenden Personen in einer großen Runde zusammensitzen oder nur ein Teil der Teilnehmer über ein Telekommunikationsnetzwerk hinzugeschaltet ist.

**[0007]** Hierbei macht sich die Erfindung zunutze, dass durch ein automatisiertes und somit objektiviertes Signalisieren sich die teilnehmenden Personen

genötigt sehen, die für eine Spracherkennung und Speicherung der erkannten Worte notwendige Sprechkultur einzuhalten. Das erfindungsgemäße Verfahren oder System hat somit den besonderen Vorteil, dass durch die Signalisierung die Teilnehmer an einer Sprachkonferenz, zumindest einige dieser Teilnehmer automatisch gezügelt werden, wenn sie anderen Teilnehmern ins Wort fallen, diese beispielsweise nicht ausreden lassen. Mittels des Systems bzw. des Verfahrens wird somit die Funktion eines Moderators implementiert.

**[0008]** Hierbei kann das Verfahren beispielsweise implementiert sein durch eine Software, die auf einen hierfür vorgesehenen Rechner abläuft. Ein derartiger Rechner kann beispielsweise innerhalb eines Telekommunikationsnetzwerkes vorhanden sein, beispielsweise in einer Vermittlungsstelle, über welche insbesondere bei einer über ein Telekommunikationsnetzwerk stattfindenden Sprachkonferenz die Konferenzschaltung schaltungstechnisch abgewickelt wird. So kann demnach automatisch beispielsweise die Spracherkennung der gesprochenen Worte und die Speicherung intern in einem Telekommunikationsnetzwerk stattfinden.

**[0009]** Gemäß einer möglichen Ausführungsvariante kann ein zeitgleiches Sprechen beispielsweise dadurch erkannt werden, dass mittels der Spracherkennung kein Wort oder kein sinnvolles Wort erkannt wird. Beispielsweise kann es vorkommen, dass bei einem zeitgleichen Sprechen bzw. bei einem zumindest teilweise zeitlichen Überlapp der von verschiedenen Teilnehmern der Sprachkonferenz gesprochenen Worte eine Spracherkennung, also beispielsweise eine Software zur Spracherkennung, die auf einem Computer implementiert ist, nicht mehr in der Lage ist, anhand seines üblicherweise vorgesehenen Sprachwortschatzes dem gesprochenen Wort ein erkanntes Wort zuzuordnen. Es findet somit keine Spracherkennung statt bzw. es wird kein sinnvolles Wort erkannt. Hierbei kann insbesondere zur Prüfung, ob es sich um ein erkanntes sinnvolles oder ein erkanntes nicht sinnvolles Wort handelt, Rückgriff genommen werden auf die bislang bei der Konferenz erkannten Worte, um sodann durch einen Vergleich des neu erkannten Wortes mit den bislang erkannten Worten auf eine Sinnkorrelation zu schließen und so festzustellen, ob es sich bei dem Wort, welches aktuell erkannt wurde, um eines handelt, welches im Sinnzusammenhang mit den vorher erkannten Worten steht oder nicht.

**[0010]** Die vorgenannten Möglichkeiten führen somit zu einer Erkennbarkeit, ob wenigstens zwei Personen gleichzeitig sprechen, um sodann in Abhängigkeit hiervon gegebenenfalls ein Signal zu generieren.

**[0011]** Wird das entsprechende Signal generiert, so

werden hierdurch die teilnehmenden Personen aufgefordert, seriell zu sprechen, wobei unter dem Begriff des seriellen Sprechens im Rahmen der Erfindung verstanden wird, dass unterschiedliche Personen nacheinander Sprachbeiträge in der Konferenz liefern.

**[0012]** In einer bevorzugten Ausgestaltung kann es auch vorgesehen sein, dass neben einer Spracherkennung zur Erkennung des gesprochenen Wortes eines Teilnehmers auch eine Sprechererkennung durchgeführt wird zur Erkennung, welcher der Teilnehmer ein Wort gesprochen hat.

**[0013]** Dies ist besonders vorteilhaft, da nach einer Speicherung der erkannten gesprochenen Worte auch nachträglich festgestellt werden kann, von welchem der Teilnehmer der entsprechende Wortbeitrag stammt. So kann bei einer derartigen Ausführung ein zeitgleiches Sprechen auch erkannt werden durch ein Erkennen von wenigstens zwei Teilnehmern, beispielsweise, die innerhalb eines vorgegebenen Zeitintervalls und insbesondere gleichzeitig gesprochen haben.

**[0014]** Hierbei kommt es nicht darauf an, dass die Worte, die die entsprechenden Teilnehmer geäußert haben, durch die Spracherkennung erkannt werden, sondern vielmehr, dass in den Äußerungen der Teilnehmer Sprachcharakteristika bzw. Sprachparameter erkannt werden, die für einen der Teilnehmer typisch sind und diesen identifizieren. So kann demnach eine Sprechererkennung, die insbesondere auch innerhalb einer Spracherkennung implementiert sein kann, selbst dann einen Teilnehmer identifizieren, wenn eine Erkennung des bzw. der gesprochenen Worte nicht möglich ist, da sich z.B. die für einen Teilnehmer typischen Sprachparameter auch aus überlagerten Sprachbeiträgen erkennen lassen.

**[0015]** So kann eine Sprechererkennung somit bevorzugt durchgeführt werden mittels wenigstens eines Sprachparameters eines Teilnehmers, der im Rahmen der Durchführung des Verfahrens bzw. dem System bekannt ist, z.B. durch eine vorherige Abspeicherung. Es kann so beispielsweise ein Vergleich stattfinden zwischen einem aktuell zu einem Wortbeitrag ermittelten Sprachparameter mit einem oder einer Gruppe von Referenzparametern, die in Verbindung mit einem bzw. bevorzugt allen Teilnehmern gespeichert sind. Durch diesen Vergleich kann somit eine Zuordnung stattfinden, welcher der Teilnehmer an der Sprachkonferenz den Sprachbeitrag geleistet hat.

**[0016]** Eine Identifizierung und somit eine Sprechererkennung kann neben dem wenigstens einen Sprachparameter, insbesondere der sich aus biometrischen Merkmalen der Sprache des Teilnehmers ableiten lässt, ergänzend auch erfolgen durch eine

Auswertung der Telekommunikationskennung des bzw. derjenigen Teilnehmer der Sprachkonferenz, von dem bzw. denen ein Sprachbeitrag kommt. Dies ist insbesondere vorteilhaft, wenn eine Sprachkonferenz mit mehreren Teilnehmern derart erfolgt, dass diese Teilnehmer über ein Telekommunikationsnetzwerk zusammengeschaltet sind.

**[0017]** Sodann ist jedem Teilnehmer bzw. jedem Telekommunikationsgerät, welches durch einen oder gegebenenfalls mehrere Teilnehmer verwendet wird, wenigstens eine Telekommunikationskennung zugeordnet. Unter einer Telekommunikationskennung können mit oder ohne einer Geräteerkennung z.B. und nicht abschließend verstanden werden: die CLI (Calling Line Identity), das HLR (Home Location Register) oder auch andere Kennungen, wie z.B. diejenigen einer SIM-Karte bei Mobiltelefonen oder sonstige Kennungen im Zusammenhang mit Telekommunikation, wie z.B. auch IP-Adresse, z.B. bei Voice-Over-IP oder auch Geräte Kennungen alleine (z.B. IMEI – International Mobile Equipment Identity, oder MAC-Adressen – Media Access Control, insbesondere bei Netzwerkkomponenten). Telekommunikationskennungen im Sinne der Erfindung können auch Kombinationen, insbesondere der vorgenannten Kennungen sein.

**[0018]** So kann durch die Auswertung dieser vorgenannten Telekommunikationskennung im Rahmen einer Sprechererkennung zusätzlich ein Teilnehmer bzw. zumindest eine Gruppe von Teilnehmern erkannt werden, sofern mehrere Teilnehmer über dasselbe Telekommunikationsgerät bzw. über denselben Telekommunikationsanschluss und somit auch über dieselbe Telekommunikationskennung an der Konferenz teilnehmen.

**[0019]** Gemäß einer weiteren bevorzugten Ausführung kann es vorgesehen sein, dass bei einer nicht erfolgten Sprechererkennung, also wenn anhand des Sprachbeitrags, z.B. eines gesprochenen Wortes nicht derjenige Teilnehmer identifiziert werden konnte, von dem dieser Sprachbeitrag kommt, dieser Teilnehmer aufgefordert wird, sich zu identifizieren.

**[0020]** Eine derartige Identifizierung kann beispielsweise durch eine Sprachaufforderung im Rahmen des Verfahrens bzw. durch das System erfolgen. Hierbei kann beispielsweise die Sprachaufforderung nur an denjenigen Teilnehmer übersendet werden, der von dem System bzw. im Rahmen des Verfahrens nicht erkannt wurde, was z.B. durch eine Telekommunikation an die Telekommunikationskennung des nicht erkannten Teilnehmers erfolgen kann.

**[0021]** Somit erhalten alle anderen Teilnehmer der Sprachkonferenz keine derartige Aufforderung bzw. nehmen diese nicht wahr, da die Aufforderung nur an den nicht erkannten Teilnehmer kommuniziert wird.

Gegebenenfalls kann es hierbei vorkommen, dass bei Teilnahme mehrerer Teilnehmer über dasselbe Telekommunikationsgerät bzw. denselben Anschluss mit nur einer Telekommunikationskennung mehrere Teilnehmer gleichzeitig die Aufforderung erhalten.

**[0022]** Nach Erhalt einer derartigen Aufforderung kann der bislang nicht identifizierte Teilnehmer beispielsweise seinen Namen nennen oder vorgegebene bzw. beliebige Worte sprechen, woraufhin wenigstens ein für die Identifizierung eines Teilnehmers nötiger Sprachparameter erfasst wird, der sodann als Referenzparameter gespeichert wird und für die zukünftige Sprechererkennung zur Verfügung steht. Auf diese Weise können bislang nicht erkannte Sprachkonferenzteilnehmer als erkannte Teilnehmer in die Sprachkonferenz integriert werden.

**[0023]** Gegebenenfalls kann es auch vorgesehen sein, bei einem bislang nicht erkannten bzw. identifizierten Teilnehmer diesen im Rahmen des Verfahrens bzw. im System zunächst als unerkannt zu führen und eine entsprechende Zuordnung zu den erkannten Worten vorzunehmen, demgemäß die gesprochenen Worte nicht einer bestimmten konkreten und gegebenenfalls namentlich bekannten Person zugeordnet werden, sondern der unerkannten Person zugeordnet werden.

**[0024]** Es kann sodann verfahrensgemäß vorgesehen sein, erst nach Beendigung einer Sprachkonferenz durch eine dann nachträglich durchzuführende Identifizierung die Wortbeiträge einer unerkannten Person der dann nachträglich identifizierten Person zuzuordnen, beispielsweise durch Änderung eines Zuordnungsparameters von dem Status „unerkannt“ auf den Status „erkannt“ bzw. den Parameter, der eine bestimmte Person identifiziert.

**[0025]** Eine Signalisierung im Sinne des erfindungsgemäßen Verfahrens kann grundsätzlich durch jede Art von Signal erfolgen, welches mit den Sinnen eines Teilnehmers erfassbar ist. Hierbei kann es bevorzugt vorgesehen sein, dass eine Signalisierung zur Aufforderung, seriell zu sprechen, z.B. durch ein optisches und/oder akustisches Signal erfolgt. Besonders bevorzugt wird es hierbei vorgesehen sein, das akustische Signal in Form einer Sprachansage erfolgen zu lassen.

**[0026]** Hierbei kann es in einer besonders bevorzugten Ausgestaltung des Verfahrens vorgesehen sein, dass eine Signalisierung nur an denjenigen Teilnehmer erfolgt, von dem eine Störung ausgeht. Beispielsweise kann die Signalisierung erfolgen an eine Telekommunikationskennung, die einem Teilnehmer zugeordnet ist. Gegebenenfalls kann der Sprachansage eine direkte Anrede namentlicher Art hinzugefügt werden, insbesondere dann, wenn eine namentliche Anrede in Verbindung mit dem vorgenannten

sprachlichen Referenzparameter zur Identifizierung eines Teilnehmers gespeichert ist.

**[0027]** Es kann sodann gezielt nur derjenige Teilnehmer die Aufforderung erhalten, von dem die Störung ausgeht, d.h. der dazwischen gesprochen oder einen anderen Teilnehmer unterbrochen hat, ohne die Aufforderung auch an die anderen Teilnehmer zu signalisieren und diese hierdurch zu stören.

**[0028]** Insbesondere zur Feststellung, welcher von zwei festgestellten Teilnehmern derjenige ist, der störend auffällt, also den Wortbeitrag eines anderen Teilnehmers durch Dazwischenreden stört oder unterbricht, kann es vorgesehen sein, die Häufigkeit einer Identifikation eines Teilnehmers zu erkannten Worten, insbesondere innerhalb eines bestimmten Zeitintervalls, zu prüfen. Wird beispielsweise festgestellt, dass innerhalb eines zurückliegenden Zeitintervalls von beispielsweise nur wenigen Sekunden, beispielsweise 10 oder 20 Sekunden oder gegebenenfalls auch einer Minute oder anderen beliebigen Zeitintervallen zu einem bestimmten Teilnehmer A eine sehr häufige, gegebenenfalls sogar ausschließlich nur für diesen Teilnehmer durchgeführte Zuordnung der erkannten Worte zu diesem speziellen Teilnehmer stattgefunden hat, so kann davon ausgegangen werden, dass bei einer sodann nur z.B. singulären oder nur mit einer geringen Häufigkeit vorgenommenen Zuordnung eines Wortbeitrags eines Teilnehmers B dieser Teilnehmer B derjenige ist, der in den Wortbeitrag des Teilnehmers A hineingesprochen und diesen somit gestört hat.

**[0029]** Ebenso ist es möglich, dass alle Teilnehmer einer Konferenz eine Signalisierung erhalten, wenn eine Spracherkennung nicht erfolgreich war, insbesondere wenn die Sprachbeiträge aller Teilnehmer mittels derselben Spracherkennung, also z.B. einer Software verarbeitet werden.

**[0030]** Gemäß dem erfindungsgemäßen Verfahren bzw. mit dem erfindungsgemäßen System kann es vorgesehen sein, dass ein erkanntes, gesprochenes Wort gespeichert wird mit einer Information, welcher der Teilnehmer das Wort gesprochen hat.

**[0031]** Dies ist insbesondere sinnvoll im Zusammenhang mit der Erstellung von Gesprächsprotokollen, so dass auch noch im Nachhinein festgestellt werden kann, von wem ein gespeichertes, gesprochenes Wort ursprünglich gekommen ist. So kann beispielsweise in Verbindung mit einem gesprochenen und erkannten Wort eine Zuordnungskennung abgespeichert werden.

**[0032]** Jeder der Teilnehmer einer Sprachkonferenz kann beispielsweise eine derartige Kennung aufweisen, die beispielsweise vorgegeben ist oder im Rahmen einer automatischen Erkennung in Verbindung

mit dem vorgenannten Referenzparameter erstellt wird. Eine derartige Zuordnungskennung kann jeglicher Art sein, beispielsweise kann es sich um einen bestimmten Inhalt eines Datensatzes handeln, z.B. um eine Zahl, eine Ziffer, den Namen, ein gesetztes oder gelöscht Bit und/oder um eine Telekommunikationskennung und/oder Gerätekennung, z.B. der jeweils vorgenannten Art.

**[0033]** Es besteht so in erfindungsgemäß besonders bevorzugter Ausgestaltung auch die Möglichkeit, dass mittels der Zuordnungskennung und/oder der Telekommunikationskennung aus den gespeicherten Worten der Teilnehmer diejenigen wenigstens eines gewünschten Teilnehmers gelöscht oder selektiert werden.

**[0034]** Beispielsweise kann hierdurch erreicht werden, dass aus der Ansammlung der gespeicherten Worte, also dem somit z. B. erstellten Protokoll einer Sprachkonferenz, diejenigen Worte erneut abgerufen oder aber auch gelöscht werden, die einem ganz bestimmten Teilnehmer zugeordnet sind.

**[0035]** Insbesondere die vorgenannten Merkmale der Zuordnung einer Kennung betreffend die Identifikation eines bestimmten Teilnehmers zu einem erkannten Wort können ganz allgemein eingesetzt werden in Verbindung mit dem vorgenannten gattungsgemäßen Verfahren zur Durchführung einer Sprachkonferenz bzw. dem vorgenannten gattungsgemäßen System, auch wenn gemäß der vorbeschriebenen Erfindung das Signalisieren einer Aufforderung zum seriellen Sprechen bei einem zeitgleichen Sprechen von wenigstens zwei Teilnehmern nicht vorgesehen ist.

**[0036]** Gemäß einer weiteren Ausführungsvariante, die in Verbindung mit allen zuvor beschriebenen Varianten einsetzbar ist, kann es vorgesehen sein, dass jedem Teilnehmer oder einer Gruppe von Teilnehmern eine eigene Spracherkennung und/oder Sprechererkennung zugeordnet ist.

**[0037]** In einer anderen Ausführung kann es auch vorgesehen sein, dass eine bezogen auf wenigstens einen, insbesondere alle Teilnehmer sprachfremde Spracherkennung eingesetzt wird. Dies führt sodann dazu, dass zwangsläufig die erkannten Worte nicht diejenigen sind, die der Teilnehmer gesprochen hat, da die gesprochenen Worte nicht im Wortschatz der sprachfremden Spracherkennung vorliegen. Dennoch werden selbst mit einer sprachfremden Spracherkennung Worte erkannt, die gespeichert, also z. B. zur Herstellung eines Protokolls mitgeschrieben werden können. Das so erstellte Protokoll bzw. der gespeicherte Satz von erkannten Worten stellt jedoch sodann keinen Klartext dar, so dass durch dieses erfindungsgemäße Verfahren eine anonymisierte Speicherung der Sprachbeiträge erfolgt.

**[0038]** Dennoch besteht die Möglichkeit, aus den gespeicherten Worten zurückzuschließen auf die tatsächlichen Wortbeiträge.

**[0039]** In einer anderen Ausgestaltung kann es auch vorgesehen sein, dass die Teilnehmer einer Sprachkonferenz in verschiedenen Sprachen kommunizieren, wobei jedem Teilnehmer bzw. einer Gruppe von Teilnehmern eine eigene Spracherkennung und/oder Sprechererkennung zugeordnet ist.

**[0040]** Eine derartige Zuordnung kann beispielsweise über die Auswertung der Telekommunikationskennung des oder derjenigen Telekommunikationsgerätes erfolgen, über das bzw. die der oder die Teilnehmer an der Sprachkonferenz teilnehmen.

**[0041]** Anhand dieser Kennung kann beispielsweise der Sprachraum ermittelt werden, d.h. die sprachliche Region, aus der die Teilnehmer stammen. So kann beispielsweise festgestellt werden, dass bei einer Telefonkonferenz mit einer CLI aus Frankreich aller Voraussicht nach die teilnehmenden Personen französisch sprechen, so dass einer französischsprachigen Person sodann eine Spracherkennung und gegebenenfalls Sprechererkennung zugeordnet wird, die auf die französische Sprache trainiert ist, also im Wesentlichen den Wortschatz der französischen Sprache umfasst.

**[0042]** Es kann sodann vorgesehen sein, dass bei der Speicherung erkannter Worte im Rahmen des erfindungsgemäßen Verfahrens die Worte der Teilnehmer in jeweils unterschiedlichen Sprachen, jedoch für die jeweilige Sprache im Klartext, abgespeichert werden. Ebenso kann es vorgesehen sein, dass in einer anderen Alternative zusätzlich zur Spracherkennung auch eine Übersetzung stattfindet, so dass trotz fremdsprachigen Wortbeitrags eine Speicherung der erkannten Wörter in einer übersetzten Sprache, beispielsweise eine für alle Teilnehmer gleiche Sprache gespeichert wird.

**[0043]** So kann beispielsweise zum Anfang einer Konferenz zwischen den Teilnehmern festgelegt werden, welches die Sprache für die Protokollierung bzw. Speicherung der Wortbeiträge sein soll, so dass sodann unabhängig von der tatsächlich verwendeten Sprache der Teilnehmer die Protokollierung nach einer Erkennung der gesprochenen Worte und gegebenenfalls Übersetzung in der gewünschten Protokollsprache gespeichert werden.

**[0044]** In einer weiteren bevorzugten Ausgestaltung des Verfahrens kann es vorgesehen sein, den gesamten Datensatz der gespeicherten erkannten und gegebenenfalls den Personen zugeordneten Worte aufzuspalten in mehrere Datensätze, wobei sodann jeder aufgespaltene Datensatz einem bestimmten Teilnehmer zugeordnet ist, so dass die einzelnen

Wortbeiträge jedes Teilnehmers separierbar sind. Gegebenenfalls können Aufspaltungen der Datensätze nicht wie vorgenannt teilnehmerbezogen, sondern bezogen auf die Telekommunikationskennung, also z.B. CLI, HLR etc. vorgenommen werden.

**[0045]** Bei einer Durchführung einer Sprachkonferenz mit mehreren fremdsprachigen Teilnehmern kann es auch vorgesehen sein, dass jeder der Teilnehmer in seiner Muttersprache spricht und weiterhin die Möglichkeit gegeben ist, die Wortbeiträge anderer Teilnehmer nach einer vorangegangenen Übersetzung, gegebenenfalls einer automatischen oder durch einen Dolmetscher erfolgten Übersetzung zu hören.

**[0046]** So kann ein Teilnehmer jeweils in seiner eigenen Muttersprache sprechen und hören, wodurch Verständigungsschwierigkeiten ausgeschlossen werden. Die Protokollierung, d.h. Speicherung der erkannten Worte kann sodann wie vorgeschrieben erfolgen, also beispielsweise im Klartext der konkreten Fremdsprache oder gegebenenfalls nach einer Übersetzung in eine gewünschte Sprache.

**[0047]** Der Ablauf des erfindungsgemäßen Verfahrens wird anhand des schematisch dargestellten Ausführungsbeispiels näher erläutert.

**[0048]** Unter Punkt 1 der Darstellung ist erkennbar, dass eine Konferenz mit den einzelnen Teilnehmern A, B und C stattfinden soll. Diese Teilnehmer kommunizieren über ein Telekommunikationsnetzwerk miteinander, was beispielsweise über die Vermittlungsstelle V erfolgen kann. Die einzelnen Telekommunikationsteilnehmer A, B und C werden so über die Vermittlungsstelle zu einer Sprachkonferenz zusammengeschaltet, wobei die Wortbeiträge von einer übergeordneten zentralen Einheit, die beispielsweise auch in der Vermittlungsstelle V realisiert sein kann oder separat im Telekommunikationsnetz angeordnet ist, erfasst werden.

**[0049]** Dies bedeutet im Zusammenhang mit diesem Ausführungsbeispiel, dass neben einer Spracherkennung zur inhaltlichen Erkennung der Worte auch eine Sprechererkennung, z.B. anhand eines oder mehrerer Sprachparameter und/oder anhand weiterer Parameter, wie die Telekommunikationsanschlusskennung erfolgt.

**[0050]** Dies erfolgt in dieser Ausführung durch die Einheit zur Sprach- bzw. Sprechererkennung SE im Schritt 2. Hierbei kann Rückgriff genommen werden auf gegebenenfalls vorhandene Datenbanken D mit Informationen über die Telekommunikationskennung bzw. die jeweils zu einem Teilnehmer gespeicherten Sprachparameter, um einen Vergleich vornehmen zu können. Ebenso können die Datenbanken D die Wortschätze gespeichert vorhalten, die der jeweils

verwendeten Spracherkennung zugrunde liegen.

**[0051]** Es kann sodann in dem weiteren 3. Verfahrensschritt SE2 festgestellt werden, ob eine erkannte sinnvolle Wortfolge vorliegt, was nur dann gegeben ist, wenn die jeweiligen Teilnehmer nacheinander sprechen, nicht jedoch, wenn sich Wortbeiträge überlagern.

**[0052]** Wird hierbei festgestellt, dass keine sinnvolle Wortfolge vorliegt, so kann im 4. Verfahrensschritt SE3 das akustische Signal an alle oder nur an den störenden Teilnehmer signalisiert werden. In der Ausführung gemäß diesem Beispiel kann es sich dabei um den gesprochenen Satz handeln „Bitte den Sprecher aussprechen lassen“.

**[0053]** Es kann sodann die weitere Verarbeitung und Speicherung des erkannten Wortes im Schritt 5 erfolgen oder gegebenenfalls auch der gestörte Teilnehmer gebeten werden, seinen Wortbeitrag zu wiederholen. Das erfindungsgemäße Verfahren wird sodann iterativ wiederholt.

### Patentansprüche

1. Verfahren zur Durchführung einer Sprachkonferenz mit mehreren Teilnehmern, die miteinander kommunizieren, insbesondere über ein Telekommunikationsnetzwerk mittels Telekommunikationsgeräten, wobei die von einem Teilnehmer gesprochenen Worte mittels einer Spracherkennung erkannt und gespeichert werden, **dadurch gekennzeichnet**, dass bei einem zeitgleichen Sprechen von wenigstens zwei Teilnehmern eine Aufforderung zum seriellen Sprechen signalisiert wird.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass ein zeitgleiches Sprechen dadurch erkannt wird, dass mittels der Spracherkennung kein Wort oder kein sinnvolles Wort erkannt wird.

3. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass neben einer Spracherkennung zur Erkennung der gesprochenen Worte auch eine Sprechererkennung durchgeführt wird zur Erkennung welcher der Teilnehmer ein Wort gesprochen hat.

4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, dass ein zeitgleiches Sprechen erkannt wird durch ein Erkennen von wenigstens zwei Teilnehmern innerhalb eines vorgegebenen Zeitintervalls, insbesondere gleichzeitig.

5. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass eine Sprechererkennung durchgeführt wird mittels wenigstens eines Sprachparameters eines Teilnehmers, insbesondere durch Vergleich mit einem oder einer Gruppe von Re-

ferenzparametern, insbesondere in Verbindung mit einer Telekommunikationskennung (CLI, HLR, SIM etc.) und/oder Geräteerkennung (IMEI, MAC etc.).

6. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass bei einer nicht erfolgten Sprechererkennung ein nicht erkannter Teilnehmer aufgefordert wird, sich zu identifizieren, insbesondere wonach wenigstens ein Referenzparameter des identifizierten Teilnehmers gespeichert wird.

7. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass eine Signalisierung zur Aufforderung seriell zu sprechen erfolgt durch ein optisches und/oder akustisches Signal, insbesondere wobei ein akustisches Signal ausgebildet ist als eine Sprachansage.

8. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass eine Signalisierung nur an denjenigen Teilnehmer erfolgt, von dem eine Störung ausgeht, insbesondere durch Signalisierung an die Telekommunikationsadresse und/oder Geräteerkennung, die einem Teilnehmer zugeordnet ist.

9. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass ein erkanntes gesprochenes Wort gespeichert wird mit einer Information, welcher der Teilnehmer das Wort gesprochen hat, insbesondere durch Abspeichern einer Zuordnungskennung in Verbindung mit einem gesprochenem Wort.

10. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass mittels der Zuordnungskennung und oder der Telekommunikationskennung aus den gespeicherten Worten der Teilnehmer diejenigen wenigstens eines gewünschten Teilnehmers gelöscht oder selektiert werden, insbesondere zum Abrufen der Worte dieses wenigstens einen Teilnehmers.

11. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass jedem Teilnehmer oder einer Gruppe von Teilnehmern eine eigene Spracherkennung und/oder Sprechererkennung zugeordnet ist.

12. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass eine bezogen auf wenigstens einen, insbesondere alle Teilnehmer sprachfremde Spracherkennung eingesetzt wird.

13. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass Teilnehmer in verschiedenen Sprachen kommunizieren, wobei die erkannten Worte in der jeweiligen Sprache und/oder in einer übersetzten Sprache, insbesondere für alle Teilnehmer gleiche Sprache gespeichert wer-

den.

14. System zur Durchführung einer Sprachkonferenz mit mehreren Teilnehmern, die miteinander kommunizieren, insbesondere über ein Telekommunikationsnetzwerk mittels Telekommunikationsgeräten, wobei die von einem Teilnehmer gesprochenen Worte mittels einer Spracherkennung erkennbar und speicherbar sind, dadurch gekennzeichnet, dass bei einem zeitgleichen Sprechen von wenigstens zwei Teilnehmern eine Aufforderung zum seriellen Sprechen signalisierbar ist.

Es folgt ein Blatt Zeichnungen

Anhängende Zeichnungen

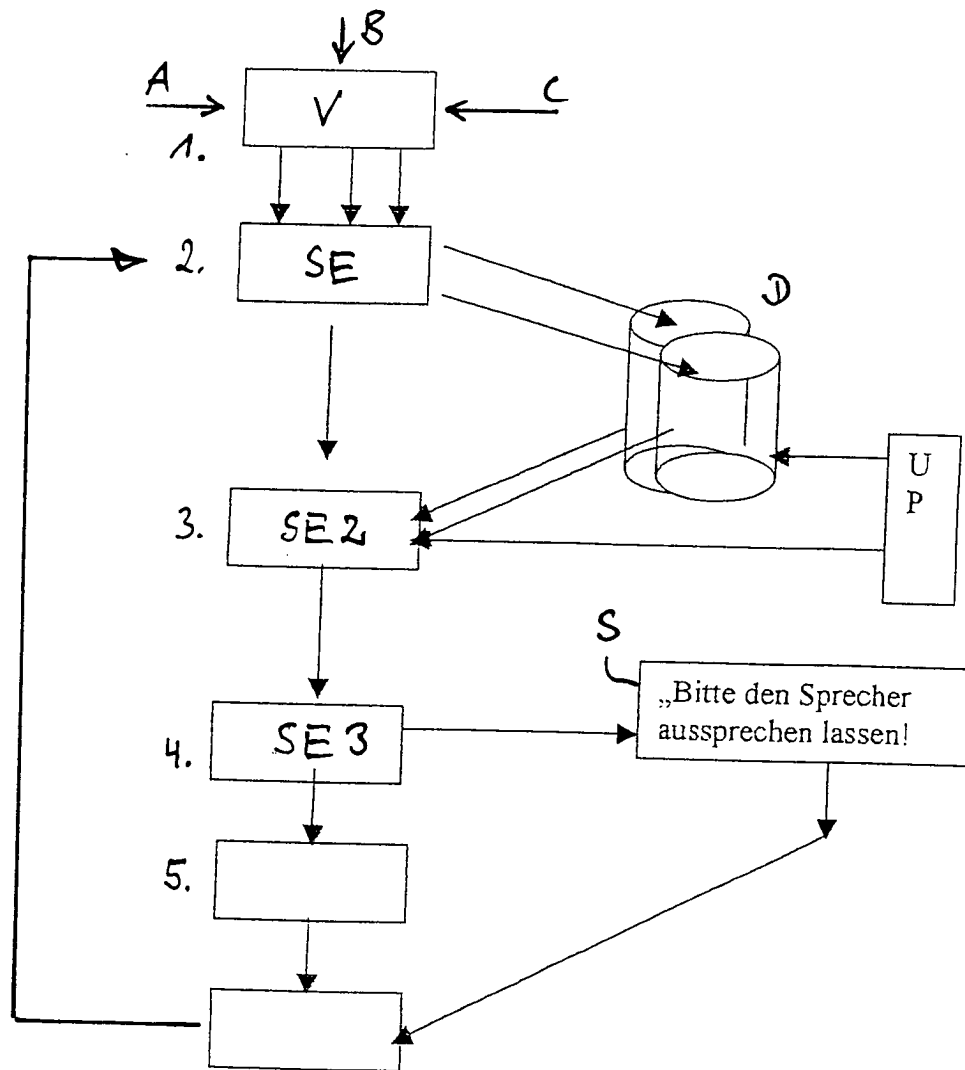


Fig. 1