



(12) **Offenlegungsschrift**

(21) Aktenzeichen: **10 2012 213 914.1**
(22) Anmeldetag: **06.08.2012**
(43) Offenlegungstag: **28.05.2014**

(51) Int Cl.: **H04M 3/493 (2006.01)**
H04M 11/00 (2006.01)

(71) Anmelder:
Reddehase, Axel, 60439, Frankfurt, DE

(72) Erfinder:
gleich Anmelder

(74) Vertreter:
Ullrich & Naumann, 69115, Heidelberg, DE

(56) Ermittelter Stand der Technik:
WO 2005/ 048 509 A2
WO 2006/ 083 690 A2

Prüfungsantrag gemäß § 44 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

(54) Bezeichnung: **Verfahren und System zum Bereitstellen einer Übersetzung eines Sprachinhalts aus einem ersten Audiosignal**

(57) Zusammenfassung: Ein Verfahren zum Bereitstellen einer Übersetzung eines Sprachinhalts aus einem ersten Audiosignal, wobei das erste Audiosignal Sprachinhalte in einer ersten Sprache enthält und über eine Sprechverbindung zwischen mindestens zwei Endgeräten (2, 3) übertragen wird, umfasst zum möglichst einfachen Unterstützen eines Gesprächsteilnehmers mit Sprachinhalten in einer zweiten Sprache die Schritte:

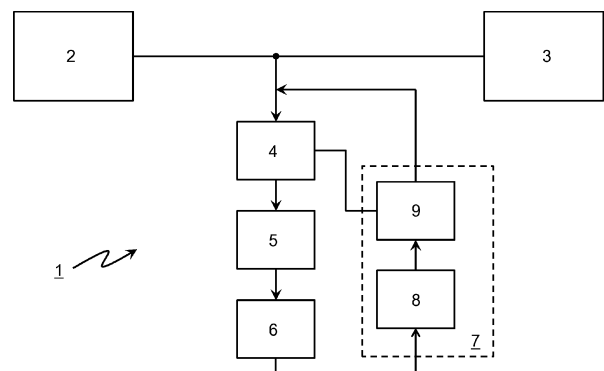
Empfangen eines Aktivierungssignals durch eine Steuereinheit (4),

Analysieren des ersten Audiosignals, das nach dem Empfang des Aktivierungssignals über die Sprechverbindung übertragen wird, mittels eines Spracherkennungssystems (5) und Erkennen eines Wortes aus dem ersten Audiosignal durch das Spracherkennungssystem (5),

Übergeben des erkannten Wortes an einen Lexikonserver (6) und Übersetzen des erkannten Wortes in eine zweite Sprache durch den Lexikonserver (6) und

Ausgeben des durch den Lexikonserver (6) übersetzten Wortes an einen Nutzer mindestens eines der Endgeräte (2, 3).

Ein entsprechendes System ist angegeben.



Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zum Bereitstellen einer Übersetzung eines Sprachinhalts aus einem ersten Audiosignal, wobei das erste Audiosignal Sprachinhalte in einer ersten Sprache enthält und über eine Sprechverbindung zwischen mindestens zwei Endgeräten übertragen wird.

[0002] Des Weiteren betrifft die Erfindung ein entsprechendes System sowie ein Kommunikationsnetzwerk, in das das System einkoppelbar ist.

[0003] Seit vielen Jahren werden intensive Bemühungen unternommen, um in einem zunehmend internationaler ablaufenden Wirtschaftsleben mit technischen Mitteln den Einfluss von Sprachbarrieren zu reduzieren. Einige Firmen haben sich das Ziel gesetzt, aus einem Audioeingangsstrom in einer ersten Sprache automatisiert einen Audioausgangsstrom in einer zweiten Sprache zu generieren. Dazu wird ein Übersetzungssystem zwischen zwei Gesprächsteilnehmern geschaltet, das die Äußerungen eines Gesprächsteilnehmers in Echtzeit in eine andere Sprache übersetzt. Wenn beispielsweise zwei Gesprächsteilnehmer, von denen der eine lediglich englisch und der andere lediglich italienisch spricht, miteinander kommunizieren möchten, so übersetzt das Übersetzungssystem die englischsprachigen Äußerungen des einen Gesprächsteilnehmers in italienisch und die italienischen Äußerungen des anderen Gesprächsteilnehmers in englisch. Auf diese Weise können – zumindest theoretisch – zwei Gesprächsteilnehmer unter Verwendung des Übersetzungssystems in Ihrer eigenen Muttersprache miteinander korrespondieren.

[0004] In der Praxis konnten bisher die mit derartigen Übersetzungssystemen verbundenen Hürden nicht überwunden werden. Eine besonders große Hürde stellt die Tatsache dar, dass jede Sprache Zweideutigkeiten aufweist, d.h. zu einer Schreibweise oder zu einer Aussprache gibt es mehrere Bedeutungen. Hinzu kommt, dass in Gesprächen gelegentlich Worte undeutlich ausgesprochen werden. Dadurch erhöht sich die Zahl möglicher Übersetzungen noch weiter. In solchen Fällen ist für die Übersetzung eines Wortes der Gesamtzusammenhang notwendig. Das Übersetzungssystem muss also den Inhalt des Gesprächs extrahieren und unter Verwendung des extrahierten Inhalts eine korrekte Übersetzung aus verschiedenen Optionen wählen. Dies ist – trotz einiger vielversprechender Ansätze – bisher noch nicht gelungen.

[0005] In der Praxis werden bisher Gespräche zwischen Gesprächsteilnehmern, die nicht über eine gemeinsame Muttersprache verfügen, in einer Sprache geführt, die die beteiligten Gesprächsteilnehmer mehr oder weniger gut beherrschen. Dabei treten im-

mer wieder Gesprächssituationen auf, in denen ein Gesprächsteilnehmer ein Wort in der genutzten Sprache nicht parat hat. Daher werden parallel zu dem Gespräch Wörterbücher genutzt, um fehlende Wörter nachzuschlagen. Dies erweist sich in der Praxis häufig als hinderlich und impraktikabel.

[0006] Der vorliegenden Erfindung liegt daher die Aufgabe zugrunde, ein Verfahren und ein System der eingangs genannten Art derart auszugestalten und weiterzubilden, dass auf möglichst einfache Art und Weise eine Unterstützung eines Gesprächsteilnehmers mit Sprachinhalten in einer zweiten Sprache möglich ist.

[0007] Erfindungsgemäß wird die voranstehende Aufgabe durch die Merkmale des Anspruchs 1 gelöst. Danach umfasst das in Rede stehende Verfahren die Schritte:

Empfangen eines Aktivierungssignals durch eine Steuereinheit,
 Analysieren des ersten Audiosignals, das nach dem Empfang des Aktivierungssignals über die Sprechverbindung übertragen wird, mittels eines Spracherkennungssystems und Erkennen eines Wortes aus dem ersten Audiosignal durch das Spracherkennungssystem,
 Übergeben des erkannten Wortes an einen Lexikonserver und Übersetzen des erkannten Wortes in eine zweite Sprache durch den Lexikonserver und
 Ausgeben des durch den Lexikonserver übersetzten Wortes an einen Nutzer mindestens eines der Endgeräte.

[0008] Hinsichtlich des Systems ist die voranstehende Aufgabe durch die Merkmale des Anspruchs 19 gelöst. Danach umfasst das System:

eine Steuereinheit, die in die Sprechverbindung einkoppelbar ist und zum Empfang eines Aktivierungssignals ausgestaltet ist,
 ein Spracherkennungssystem, das kommunizierend mit der Steuereinheit verbunden ist und das zur Analyse des von der Steuereinheit empfangenen ersten Audiosignals und zum Erkennen eines in dem Audiosignal enthaltenen Wortes in einer ersten Sprache ausgestaltet ist,
 ein Lexikonserver, der kommunizierend mit dem Spracherkennungssystem verbunden ist und der zum Übersetzen des durch das Spracherkennungssystem erkannten Wortes in ein Wort der zweiten Sprache ausgestaltet ist,
 einem Ausgabesystem, das kommunizierend mit dem Lexikonserver verbunden ist und zum Ausgeben des durch den Lexikonserver übersetzten Wortes an einen Nutzer mindestens eines der Endgeräte ausgestaltet ist.

[0009] In erfindungsgemäßer Weise ist zunächst erkannt worden, dass Teile der aus der Praxis bekannten Lösungsansätze auf verblüffend einfache Art

und Weise miteinander kombiniert werden können. Erfindungsgemäß ist nämlich erkannt worden, dass es in der Praxis selten notwendig ist, ein komplettes Gespräch zu übersetzen. Vielmehr verfügen Gesprächsteilnehmer häufig über ausreichende Kenntnisse in einer gemeinsamen Sprache, in der ein Gespräch stattfinden soll und benötigen nur bei wenigen Worten Unterstützung. Zudem wird in gewöhnlichen Gesprächssituationen, die in der gemeinsamen Sprache geführt werden, selten eine Unterstützung bei solchen Wörtern benötigt, die eine sehr große Anzahl von möglichen Bedeutungsinhalten aufweisen. Worte mit vielen Bedeutungen sind üblicherweise recht einfache Worte, die ein Gesprächsteilnehmer häufig kennt und aktiv nutzen kann. Vielmehr besteht häufig bei spezielleren Wörtern Unterstützungsbedarf, wobei diese Worte meist lediglich eine oder wenige Bedeutungsinhalte haben. Dadurch wird es möglich, ein Übersetzungssystem zu implementieren, das Übersetzungen eines Wortes oder einer Redewendung bereitstellt.

[0010] Bei dem erfindungsgemäßen Verfahren und dem erfindungsgemäßen System wird über die Sprechverbindung zwischen mindestens zwei Endgeräten ein erstes Audiosignal ausgetauscht. Das erste Audiosignal bezeichnet die Audioinformationen, die üblicherweise von jedem der an der Sprachverbindung beteiligten Endgeräte ausgesendet werden kann. In dem ersten Audiosignal sind Sprachinhalte eines Teilnehmers an einem Endgerät in einer ersten Sprache enthalten. Da das erfindungsgemäße System „wissen“ muss, wann eine Übersetzung des ersten Audiosignals gewünscht ist, wird erfindungsgemäß ein Aktivierungssignal verwendet, das in die Sprechverbindung gesendet und von einer Steuereinheit empfangen wird. Über das Aktivierungssignal wird der Steuereinheit signalisiert, dass nachfolgend in dem ersten Audiosignal ein zu übersetzendes Wort oder eine zu übersetzende Wortfolge enthalten sein wird.

[0011] Die Steuereinheit ist vorzugsweise im Netzwerk zwischen den beiden Gesprächsteilnehmern implementiert. Das Segment des ersten Audiosignals, das nach dem Empfang des Aktivierungssignals über die Sprechverbindung übertragen wird, wird mittels eines Spracherkennungssystems analysiert. Wenn das Spracherkennungssystem aus dem Segment des ersten Audiosignals ein Wort erkennt, wird das erkannte Wort an einen Lexikonserver übergeben, der das erkannte Wort in eine zweite Sprache übersetzt. Das durch den Lexikonserver übersetzte Wort wird mittels eines Ausgabesystems an einen der Nutzer mindestens eines der Endgeräte ausgegeben. Dabei kann die Ausgabe derart erfolgen, dass eine möglichst geringe Beeinträchtigung des „normalen“ Gesprächsflusses zwischen den Endgeräten erreicht wird.

[0012] Das erfindungsgemäße Verfahren und das erfindungsgemäße System kann im Zusammenhang mit einer Vielzahl von Endgeräten eingesetzt werden. Klassische Telefone lassen sich ebenso verwenden wie IP (Internet Protocol)-basierte Endgeräte. Als echte Hardware vorhandene Endgeräte sind ebenso nutzbar wie softwarebasierte Telefonielösungen. Wichtig ist lediglich, dass zwischen mindestens zwei Endgeräten eine Sprechverbindung zum Austausch von Sprachinhalten aufgebaut werden kann. Entsprechend ist es für das erfindungsgemäße Verfahren und das erfindungsgemäße System unerheblich, auf welche Weise die Sprechverbindung aufgebaut und betrieben wird. Verbindungsorientierte Systeme sind ebenso einsetzbar wie paketorientierte Systeme. Das Audiosignal kann sowohl als analoges als auch als digitales Signal übertragen werden.

[0013] Für das Übertragen des Aktivierungssignals können verschiedene Techniken verwendet werden. So ist es denkbar, ein Tonsignal zu nutzen, das von der Steuereinheit als Aktivierungssignal interpretiert wird. Eine mögliche Alternative besteht in der Nutzung eines Steuerkanals, wie er beispielsweise bei ISDN-Systemen vorhanden ist. Bei IP-basierten Systemen kann das Aktivierungssignal als spezielles Datenpaket oder als Flag in einem Datenpaket an die Steuereinheit übertragen werden.

[0014] In einer Ausgestaltung der Erfindung wird der Schritt des Ausgebens des durch den Lexikonserver übersetzten Wortes derart durchgeführt, dass das übersetzte Wort mittels eines Text-zu-Sprache-Wandlers in ein zweites Audiosignal umgewandelt und das derart erzeugte zweite Audiosignal an den Nutzer ausgegeben wird. Damit ist es auf einfache Art und Weise möglich, einen Gesprächsteilnehmer mit Übersetzungshilfen in der zweiten Sprache zu versorgen.

[0015] Bevorzugter Weise erfolgt die Ausgabe des zweiten Audiosignals in die Sprechverbindung. Auf diese Weise kann das zweite Audiosignal sehr gut in den normalen Gesprächsfluss über die Sprechverbindung integriert werden, ohne dass es durch Nutzer negativ wahrgenommen wird.

[0016] In einer anderen Ausgestaltung der Erfindung wird der Schritt des Ausgebens des durch den Lexikonserver übersetzten Wortes durch eine Wiedergabe in Form von Text realisiert. Das übersetzte Wort wird an den Nutzer übertragen und bei dem Nutzer dargestellt. Vorzugsweise erfolgt dies über eine Webseite. Ein Nutzer würde hierbei mit einem Webbrowser eine Webseite besuchen und sich gegebenenfalls dort anmelden und/oder auf andere Weise authentifizieren. Danach würde – eventuell nach weiteren Schritten, wie beispielsweise eine Registrierung der Gesprächsverbindung – ein in dem ersten Audiosignal erkanntes und durch den Lexikonserver über-

setzes Wort durch einen Webserver auf der besuchten Webseite dargestellt. Auch damit kann dem Nutzer in Echtzeit und auf einfache Art und Weise eine Übersetzung eines Sprachinhalts aus dem ersten Audiosignal wiedergegeben werden.

[0017] In einer bevorzugten Ausgestaltung wird das Aktivierungssignal, das durch das Steuergerät empfangen wird, durch eines der Endgeräte erzeugt. Allerdings kann das Aktivierungssignal auch durch ein anderes Gerät erzeugt werden, auf das ein Nutzer eines Endgeräts während des Bestehens der Sprechverbindung Zugriff hat. So könnte sich ein Benutzer auf einer Webseite einloggen und könnte dort durch Klicken eines Buttons das Erzeugen eines Aktivierungssignals auslösen. Wesentlich ist lediglich, dass einem Nutzer Mittel zur Verfügung gestellt sind, mit denen das Erzeugen eines Aktivierungssignals initiiert werden kann.

[0018] In einer bevorzugten Ausgestaltung der Erfindung wird das Erzeugen eines Aktivierungssignals durch Drücken einer Taste an dem Endgerät ausgelöst. Der Begriff „Taste“ ist weit zu verstehen. Ein mechanisch drückbarer Knopf im klassischen Sinne kann ebenso eine „Taste“ bezeichnen wie eine Tastfläche auf einem berührungssensitiven Display. Durch Betätigen der Taste wird das Endgerät dazu veranlasst, ein Aktivierungssignal zu erzeugen und an die Steuereinheit zu übertragen.

[0019] In einer anderen bevorzugten Ausgestaltung wird das Erzeugen eines Aktivierungssignals durch Erkennen eines Schlüsselwortes in dem ersten Audiosignal ausgelöst. Zur Vermeidung eines versehentlichen Auslösens, wird vorzugsweise ein Schlüsselwort gewählt, das üblicherweise in einem gewöhnlichen Gesprächsfluss nicht vorkommt. Dabei wäre es auch denkbar, als „Schlüsselwort“ ein Geräusch oder eine wie auch immer geartete Lautfolge zu wählen.

[0020] Zum Erkennen des Schlüsselwortes ist vorzugsweise bei dem Endgerät eine Aktivierungseinheit vorgesehen. Die Aktivierungseinheit analysiert während einer bestehenden Sprechverbindung fortwährend das erste Audiosignal, das an dem Endgerät erzeugt und in die Sprechverbindung gesendet wird. Da die Aktivierungseinheit lediglich das Vorkommen des Schlüsselwortes in dem ersten Audiosignal erkennen muss, vereinfacht sich die Analyse des ersten Audiosignal erheblich. Auf diese Weise wird die Aktivierungseinheit relativ wenig belastet. Das Schlüsselwort kann in einem Speicher bei der Aktivierungseinheit hinterlegt und durch den Nutzer änderbar sein. Bei Erkennen des Schlüsselwortes in dem ersten Audiosignal sendet die Aktivierungseinheit das Aktivierungssignal an die Steuereinheit oder veranlasst das Endgerät zum Aussenden des Aktivierungssignals.

[0021] In einer weiteren Ausgestaltung der Erfindung wird das Erzeugen eines Aktivierungssignals durch mechanische Bewegung des Endgeräts ausgelöst. Viele der heute weitverbreiteten Smartphones sind mit einem Beschleunigungssensor ausgestattet, der zum Auslösen des Erzeugens eines Aktivierungssignals genutzt werden kann. Wird beispielsweise die Neigung des Endgeräts auf definierte Art und Weise geändert, kann das Erzeugen eines Aktivierungssignals ausgelöst werden. Zudem kann das Endgerät beispielsweise durch Schütteln zum Erzeugen eines Aktivierungssignals veranlasst werden.

[0022] Verschiedene der zuvor genannten Aktivierungsereignisse können auch kombiniert werden.

[0023] Wenn verschiedene Aktivierungsereignisse ermöglicht sind, kann durch Wahl eines der mehreren Aktivierungsereignisse eine von mehreren zweiten Sprachen gewählt werden. So wäre denkbar, dass bei Drücken der Sterntaste die zweite Sprache als englisch, bei Drücken der Rautetaste die zweite Sprache als spanisch und bei Drücken der Null die zweite Sprache als deutsch festgelegt wird. Je nach derart definierter zweiter Sprache würde der Lexikonserver das in dem ersten Audiosignal erkannte Wort in eine andere zweite Sprache übersetzen.

[0024] In einer Ausgestaltung des Verfahrens bzw. des Systems wird die Sprechverbindung nach Empfang eines Aktivierungssignals bei der Steuereinheit unbeeinflusst weitergeführt. Die an der Sprechverbindung teilnehmenden weiteren Teilnehmer nehmen damit wahr, wenn nach einem Aktivierungssignal das zu übersetzende Wort als erstes Audiosignal in die Sprechverbindung gesendet wird.

[0025] In einer alternativen Ausgestaltung wird nach Empfangen des Aktivierungssignals das erste Audiosignal unterdrückt, das von dem das Aktivierungssignal aussendenden Endgerät in die Sprechverbindung übertragen wird. Auf diese Weise kann gegenüber den anderen an der Sprechverbindung beteiligten Teilnehmern die Verwendung des Übersetzungssystems verborgen bleiben. Die anderen Teilnehmer nehmen lediglich ein kurzes Schweigen wahr, was in üblichen Gesprächen gelegentlich auftauchen kann und nicht ungewöhnlich auffällt.

[0026] Entsprechend könnte – wenn die Ausgabe des übersetzten Wortes unter Verwendung eines zweiten Audiosignals erfolgt – bei Ausgeben des zweiten Audiosignals in die Sprechverbindung das zweite Audiosignal an sämtliche an der Sprechverbindung beteiligten Endgeräte übertragen werden. Alternativ könnte das zweite Audiosignal lediglich an das Endgerät übertragen werden, das das Aktivierungssignal ausgesendet hat. In beiden Fällen würde das zweite Audiosignal mit dem ersten Audiosignal gemischt und sowohl das erste Audiosignal als

auch das zweite Audiosignal über die Sprechverbindung übertragen werden. Auf diese Weise kann beispielsweise, wenn das zweite Audiosignal lediglich an das das Aktivierungssignal aussendende Endgerät übertragen wird, der betreffende Gesprächsteilnehmer sowohl das von einem anderen Gesprächsteilnehmer empfangene erste Audiosignal als auch das übersetzte Wort in der zweiten Sprache hören.

[0027] Zur Erzeugung des zweiten Audiosignals durch den Text-zu-Sprache-Wandler kann der Wandler in einer besonders einfachen Ausgestaltung auf eine Computergenerierten Stimme zurückgegriffen werden. Alternativ kann ein Nutzer eine „Stimmprobe“ hinterlegen, basierend auf der der Text-zu-Sprache-Wandler das zweite Audiosignal erzeugt. Damit lässt sich ein zweites Audiosignal erzeugen, das einem durch den Nutzer des Endgeräts erzeugten Audiosignal verblüffend ähnlich klingt. Auf diese Weise kann, wenn das zweite Audiosignal an sämtliche Teilnehmer der Sprechverbindung übertragen wird, den weiteren Gesprächsteilnehmern suggeriert werden, dass der Nutzer des das Aktivierungsereignis auslösenden Endgeräts selbst das übersetzte Wort gesprochen hätte. Auf diese Weise kann die Integration des erfindungsgemäßen System und des erfindungsgemäßen Verfahrens in den üblichen Gesprächsfluss noch weiter verbessert werden.

[0028] Bei dem Übersetzen eines durch das Spracherkennungssystem erkannten Wortes kann der Lexikonserver auf eine Wörterbuchdatenbank zurückgreifen. Die Wörterbuchdatenbank kann als Teil des Lexikonserver implementiert sein. Alternativ lässt sich eine Datenbank auf einem entfernt von dem Lexikonserver angeordneten Datenbankserver nutzen. Auf diese Weise können äußerst flexibel verschiedene Wissensbasen integriert werden. Dabei ist es auch möglich, dass der Lexikonserver ein Basiswörterbuch lokal gespeichert hat, während besondere Bedeutungsinhalte über einen entfernt angeordneten Datenbankserver eingebunden werden.

[0029] Zur weiteren Verbesserung der Übersetzung eines erkannten Wortes kann die Wörterbuchdatenbank in Abhängigkeit des zu erwartenden Gesprächsinhalts, der über die Sprechverbindung sehr wahrscheinlich übertragen wird, angepasst werden. So werden beispielsweise bei einem Gespräch zwischen Medizinern andere Worte verwendet und potenziell erfragt werden als bei einem Gespräch zwischen Finanzfachleuten oder Anwälten. Der erwartete Gesprächsinhalt ließe sich dabei auf verschiedenste Art und Weise bekanntgeben. So wäre eine über das Endgerät menügeführte Einrichtung ebenso denkbar wie das Einloggen auf einer Webseite, über die ein Gesprächsteilnehmer die jeweiligen Wörterbücher einbuht.

[0030] Bevorzugter Weise wird der Schritt des Analysierens des ersten Audiosignal durch das Spracherkennungssystem durch den Empfang des Aktivierungssignals gestartet und durch den Empfang eines Terminierungssignals beendet. Auf diese Weise lässt sich klar abgrenzen, welcher Teil des ersten Audiosignals zu übersetzen ist. Das Terminierungssignal kann – ähnlich wie das Aktivierungssignal – auf verschiedenste Art und Weise gebildet sein. So kann beispielsweise das Erkennen einer Wortlücke das Erzeugen eines Terminierungssignals auslösen. Dadurch kann das Erzeugen eines Terminierungssignals besonders einfach bei solchen Systemen realisiert werden, bei denen lediglich einzelne Worte übersetzt werden müssen. Das Erkennen von Wortlücken ist in der Praxis der Spracherkennungssysteme hinlänglich bekannt.

[0031] Alternativ kann das Terminierungssignal durch den Ablauf einer vorgebbaren Zeitspanne bewirkt werden. Auf diese Weise lassen sich Wortgruppen an das Übersetzungssystem übermitteln, so dass auch kurze Redewendungen übersetzbar sind. Die Tatsache, dass eventuell einzelne, nicht zur Übersetzung vorgesehene Worte mit an das Spracherkennungssystem übermittelt werden, wird zugunsten einer besonders einfachen Implementierung und einer besonders einfachen Bedienung billigend in Kauf genommen.

[0032] Alternativ kann das Erzeugen eines Terminierungssignals durch Drücken einer Taste an dem das Aktivierungssignal erzeugenden Endgeräts ausgelöst werden. Die Taste kann dabei eine der Tasten sein, die das Erzeugen eines Aktivierungssignals hervorrufen können. Wenn beispielsweise an einem Endgerät die Tasten *, # und 0 das Erzeugen eines Aktivierungssignals auslösen, kann nach Aussenden des Aktivierungssignals das erneute Drücken einer dieser Tasten das Generieren eines Terminierungssignals hervorrufen. Alternativ können das Erzeugen eines Aktivierungssignals und das Erzeugen eines Terminierungssignals unterschiedlichen Tasten zugeordnet sein. So kann beispielsweise ein Aktivierungssignal durch Drücken der Raute- und der Stern-taste erzeugt werden, während das Drücken der 0 ein Terminierungssignal erzeugt.

[0033] Ergänzend oder alternativ kann das Erkennen eines Schlüsselworts an dem das Aktivierungssignal auslösenden Endgeräts das Erzeugen eines Terminierungssignals auslösen. Hierbei kann das gleiche Schlüsselwort wie zum Erzeugen des Aktivierungssignals verwendet werden. Alternativ kann für das Erzeugen eines Terminierungssignals ein eigenes Schlüsselwort Verwendung finden. Der Begriff „Schlüsselwort“ ist ebenso wie bei der Erzeugung eines Aktivierungssignals sehr allgemein zu verstehen.

[0034] Alternativ kann das Erzeugen eines Terminierungssignals durch mechanische Bewegung des das Aktivierungssignal erzeugenden Endgerätes auslöset werden. Auch hier ist ein Schütteln des Endgerätes ebenso denkbar wie ein mechanisches Verkippen des Endgerätes.

[0035] Es sei noch darauf hingewiesen, dass die Erfindung nicht auf die Verwendung mit zwei unterschiedlichen Sprachen beschränkt ist. Vielmehr können mit dem erfindungsgemäßen Verfahren und dem erfindungsgemäßen System auch Übersetzungen von Begriffen, wie Fachwörtern oder wenig gebräuchlichen Wörtern, angeboten werden. Dadurch sind das erfindungsgemäße System und das erfindungsgemäße Verfahren besonders universell einsetzbar.

[0036] Es gibt nun verschiedene Möglichkeiten, die Lehre der vorliegenden Erfindung in vorteilhafter Weise auszugestalten und weiterzubilden. Dazu ist einerseits auf die dem Anspruch 1 bzw. 19 nachgeordneten Ansprüche und andererseits auf die nachfolgende Erläuterung von bevorzugten Ausführungsbeispielen der Erfindung anhand der Zeichnung zu verweisen. In Verbindung mit der Erläuterung der bevorzugten Ausführungsbeispiele der Erfindung anhand der Zeichnung werden auch im Allgemeinen bevorzugte Ausgestaltungen und Weiterbildungen der Lehre erläutert.

[0037] In der Zeichnung zeigen

[0038] Fig. 1 ein Blockschaltbild eines ersten Ausführungsbeispiels unter Verwendung eines Text-zu-Sprache-Wandlers und eines Übertragungssystems als Ausgabesystem und

[0039] Fig. 2 ein Blockschaltbild eines zweiten Ausführungsbeispiels unter Verwendung eines Webserver als Ausgabesystem.

[0040] Fig. 1 zeigt ein Blockschaltbild eines ersten Ausführungsbeispiels. Das erfindungsgemäße System 1 ist an eine Sprechverbindung zwischen zwei Endgeräten 2 und 3 ankoppelbar. Die Endgeräte 2 und 3 sind in diesem Ausführungsbeispiel durch gewöhnliche Telefone implementiert, die untereinander eine Sprechverbindung über ein analoges oder digitales Telefonnetzwerk aufbauen. Das System 1 ist sowohl entfernt von dem Endgerät 2 als auch entfernt vom Endgerät 3 realisiert. Das System 1 kann als Mehrwertdienst durch den Telefonnetzbetreiber implementiert sein oder kann als weiterer Gesprächsteilnehmer ähnlich einer Konferenzschaltung integriert werden. Alternativ kann die Ankoppelbarkeit auch durch ein Interface an dem jeweiligen Endgerät realisiert sein, durch das Sprechverbindungen zu anderen Endgeräten aufgebaut werden und das zur Durchführung des erfindungsgemäßen Verfah-

rens auf das entfernt angeordnete erfindungsgemäße System 1 zugreift.

[0041] Das System 1 besteht aus einer Steuereinheit 4, einem Spracherkennungssystem 5, einem Lexikonserver 6 und einem Ausgabesystem 7, wobei das Ausgabesystem 7 durch einen Text-zu-Sprache-Wandler 8 und ein Übertragungssystem 9 gebildet ist. Die Steuereinheit 4 ist an die Sprechverbindung zwischen dem Endgerät 2 und 3 ankoppelbar, so dass die Steuereinheit 4 das von dem Endgerät 2 zum Endgerät 3 übertragene erste Audiosignal oder das von dem Endgerät 3 zum Endgerät 2 übertragene erste Audiosignal empfangen kann. Die Steuereinheit 4 ist kommunizierend mit dem Spracherkennungssystem 5 verbunden und kann empfangene Segmente des ersten Audiosignals für eine Analyse an das Spracherkennungssystem 5 weiterleiten. Das Spracherkennungssystem 5 ist kommunizierend mit dem Lexikonserver 6 verbunden, wobei über diese Verbindung ein aus dem ersten Audiosignal erkanntes Wort an den Lexikonserver 6 übertragen wird. Der Lexikonserver 6 ist kommunizierend mit dem Text-zu-Sprache-Wandler 8 verbunden. Über diese Verbindung wird das übersetzte Wort an den Text-zu-Sprache-Wandler 8 übermittelt. Dort wird das übersetzte Wort in ein zweites Audiosignal gewandelt. Das durch den Text-zu-Sprache-Wandler 8 erzeugte zweite Audiosignal wird über eine Verbindung zwischen dem Text-zu-Sprache-Wandler 8 an das Übertragungssystem 9 übertragen.

[0042] Das Übertragungssystem 9 ist wiederum mit der Sprechverbindung zwischen dem Endgerät 2 und 3 verbunden, um das zweite Audiosignal in die Sprechverbindung zu übertragen. Es ist eine Aufgabe des Übertragungssystems 9 das erste Audiosignal mit dem zweiten Audiosignal zu mischen. Hierzu empfängt das Übertragungssystem 9 von der Steuereinheit 4 zusätzlich das erste Audiosignal und mischt die beiden Audiosignale zu einem gemeinsamen Audiosignal. Dieses Audiosignal wird an die Sprechverbindung zwischen dem Endgerät 2 und 3 übertragen. Zusätzlich können die Steuereinheit 4 Steuerinformationen an das Übertragungssystem 9 übertragen, die beispielsweise definieren, wie die Pegel des ersten und zweiten Audiosignals zueinander eingestellt werden oder ob das zweite Audiosignal dem Endgerät 2, dem Endgerät 3 oder beiden Endgeräten zur Verfügung gestellt werden soll.

[0043] Beim Betrieb des erfindungsgemäßen Systems 1 empfängt die Steuereinheit 4 ein Aktivierungssignal. Ohne Beschränkung des Ausführungsbeispiels wird nachfolgend davon ausgegangen, dass das Aktivierungssignal durch das Endgerät 2 gesendet wird. Nachdem die Steuereinheit 4 das Aktivierungssignal von dem Endgerät 2 empfangen hat, zeichnet die Steuereinheit 4 das erste Audiosignal auf, das von Endgerät 2 an das Endgerät 3 übertra-

gen wird. Abhängig von der Konfiguration des Systems veranlasst die Steuereinheit 4, dass das erste Audiosignal des Endgeräts 2 nicht an das Endgerät 3 gesendet wird oder dass die Gesprächsverbindung unverändert weiterläuft.

[0044] Sobald die Steuereinheit 4 ein Terminierungssignal empfängt, das sinnvoller Weise von dem gleichen Gerät gesendet wird wie das Aktivierungssignal (hier Endgerät 2), wird die Aufzeichnung beendet und gegebenenfalls das erste Audiosignal des Endgeräts 2 wieder „normal“ über die Sprechverbindung übertragen. Das derart generierte Segment des ersten Audiosignals wird an das Spracherkennungssystem 5 übertragen. Dort wird das Segment analysiert und eine Spracherkennung durchgeführt. Hierbei lassen sich verschiedenste, aus der Praxis bekannte Spracherkennungssysteme einsetzen.

[0045] Das durch das Spracherkennungssystem 5 erkannte Wort oder die durch das Spracherkennungssystem 5 erkannte Wortfolge wird nachfolgend an den Lexikonserver 6 übertragen, der das erkannte Wort oder die erkannte Wortfolge in einer Wörterbuchdatenbank sucht. Hierbei können auch Konflikte, in denen mehrere alternative Bedeutungen gefunden werden, aufgelöst werden. So könnte nicht lediglich ein übersetztes Wort oder eine übersetzte Wortfolge an den Text-zu-Sprache-Wandler 8 übermittelt werden, sondern eine Gruppe von übersetzten Worten oder Wortfolgen. Der Text-zu-Sprache-Wandler 8 kann also die verschiedenen Alternativen zusammen erhalten.

[0046] Der Text-zu-Sprache-Wandler 8 wandelt das übersetzte Wort oder die übersetzte Wortfolge in ein zweites Audiosignal um. Text-zu-Sprache-Wandler sind aus dem Stand der Technik hinlänglich bekannt. Bei der Wandlung des Textes in Sprache wird auf Sprachsynthese Datenbanken zurückgegriffen, die auch Informationen über die Sprache des Nutzers des Endgeräts 2 enthalten kann.

[0047] Das durch den Text-zu-Sprache-Wandler 8 erzeugte zweite Audiosignal wird an das Übertragungssystem 9 übergeben, das zusätzlich von der Steuereinheit 4 das erste Audiosignal erhält. Das Übertragungssystem 9 mischt das erste Audiosignal und das zweite Audiosignal derart zusammen, dass sich der Pegel des gemischten Audiosignals im Vergleich zu dem ersten Audiosignal nicht erhöht. Das Mischen und/ oder Einkopplern in die Sprechverbindung erfolgt dabei derart, dass Rückkopplungen des ersten Audiosignals vermieden werden. Das gemischte Audiosignal wird – abhängig von der Konfiguration, die das Übertragungssystem 9 von der Steuereinheit 4 erhält – dem Endgerät 2 oder beiden an der Sprechverbindung beteiligten Endgeräten 2 und 3 übermittelt.

[0048] Fig. 2 zeigt ein zweites Ausführungsbeispiel des erfindungsgemäßen Systems 1'. System 1' ist im Wesentlichen identisch zu dem System 1 des ersten Ausführungsbeispiels. Allerdings ist das Ausgabesystem 7 in diesem Ausführungsbeispiel durch einen Webserver 10 gebildet, der kommunizierend mit dem Lexikonserver 6 verbunden ist. Der Webserver 10 ist zum Generieren einer Webseite ausgestaltet, die an einer Anzeigevorrichtung 11 bei dem Nutzer eines Endgeräts – entsprechend Fig. 2 der Nutzer des Endgeräts 2 – dargestellt werden kann. Die Webseite wird über eine Internetverbindung 12 von dem Webserver 10 in die Anzeigevorrichtung 11 geladen. Unter dem Begriff Anzeigevorrichtung sind sämtliche Geräte zu verstehen, die in der Lage sind Webinhalte aus dem Internet zu laden und auf einer Anzeige darzustellen. Anzeigevorrichtungen können also durch Personal Computer, Laptops, Tablet-Rechner, Smartphones also auch durch Set-Top-Boxen für Fernsehgeräte gebildet sein.

[0049] Beim Betrieb des Systems 1' empfängt die Steuereinheit 4 ein Aktivierungssignal. Ähnlich wie bei Fig. 1 wird – ohne Beschränkung der Allgemeinheit – davon ausgegangen, dass Endgerät 2 das Aktivierungssignal aussendet. Daher ist lediglich bei Endgerät 2 eine Anzeigevorrichtung 11 eingezeichnet. Eine entsprechende Einrichtung jedoch auch bei Endgerät 3 vorgesehen sein. Bei diesem Ausführungsbeispiel wäre auch denkbar, das Aktivierungssignal von der Anzeigevorrichtung 11 aus an das System 1' zu senden.

[0050] Nachdem die Steuereinheit 4 das Aktivierungssignal empfangen hat, zeichnet die Steuereinheit 4 das erste Audiosignal auf, das von Endgerät 2 an Endgerät 3 übertragen wird. Abhängig von der Konfiguration des Systems 1' veranlasst die Steuereinheit 4, dass das erste Audiosignal des Endgeräts 2 nicht an das Endgerät 3 gesendet wird oder dass die Gesprächsverbindung unverändert weiterläuft.

[0051] Sobald die Steuereinheit 4 ein Terminierungssignal vom Endgerät 2 (oder auch von der Anzeigevorrichtung 11) empfängt, das durch Drücken einer Taste oder eines Buttons an dem Endgerät 2 ausgelöst wird, wird die Aufzeichnung gestoppt und gegebenenfalls das erste Audiosignal des Endgeräts 2 wieder „normal“ über die Sprechverbindung übertragen. Das derart generierte Segment des ersten Audiosignals wird an das Spracherkennungssystem 5 übertragen. Dort wird das Segment analysiert und eine Spracherkennung durchgeführt.

[0052] Das durch das Spracherkennungssystem 5 erkannte Wort oder Wortfolge wird an den Lexikonserver 6 übertragen, der das erkannte Wort oder die erkannte Wortfolge in einer Wörterbuchdatenbank sucht. Die aus der Wörterbuchdatenbank bezogenen Worte, Wortfolgen oder Gruppen von Worten oder

Wortfolgen werden an den Webserver **10** übergeben. Der Webserver erzeugt eine Webseite, die das/die von dem Lexikonserver **6** übersetzte/n Wort/e enthält, und überträgt die Webseite über die Internetverbindung **12** an die Anzeigevorrichtung **11** bei dem Nutzer des Endgeräts **2** bzw. aktualisiert die von der Anzeigevorrichtung **11** geladene Seite. Mechanismen zur Implementierung dieser Funktionen sind aus der Praxis bekannt.

[0053] In Ergebnis kann ein Nutzer durch Drücken einer vordefinierten Taste, beispielsweise die Raute-taste, das System **1'** dazu veranlassen, ein nach dem Drücken der Taste über die Sprechverbindung übertragenes Wort zu erkennen, in eine andere Sprache zu übersetzen und an einer Anzeigevorrichtung **11** bei dem Nutzer darzustellen. Da diese Mechanismen sehr schnell ablaufen können, erhält der Nutzer mit nur geringer Verzögerung Übersetzungshilfen auf seiner Anzeigevorrichtung und kann optimal in einem Gespräch unterstützt werden.

[0054] Hinsichtlich weiterer vorteilhafter Ausgestaltungen des erfindungsgemäßen Verfahrens und des erfindungsgemäßen Systems wird zur Vermeidung von Wiederholungen auf den allgemeinen Teil der Beschreibung sowie auf die beigefügten Ansprüche verwiesen.

[0055] Schließlich sei ausdrücklich darauf hingewiesen, dass die voranstehend beschriebenen Ausführungsbeispiele der erfindungsgemäßen Vorrichtung lediglich zur Erörterung der beanspruchten Lehre dienen, diese jedoch nicht auf die Ausführungsbeispiele einschränken. Des Weiteren sei darauf hingewiesen, dass Aspekte der beiden Ausführungsbeispiele ohne Weiteres kombiniert werden können.

Bezugszeichenliste

1	System
2	Endgerät
3	Endgerät
4	Steuereinheit
5	Spracherkennungssystem
6	Lexikonserver
7	Ausgabesystem
8	Text-zu-Sprache-Wandler
9	Übertragungssystem
10	Webserver
11	Anzeigevorrichtung
12	Internetverbindung

Patentansprüche

1. Verfahren zum Bereitstellen einer Übersetzung eines Sprachinhalts aus einem ersten Audiosignal, wobei das erste Audiosignal Sprachinhalte in einer ersten Sprache enthält und über eine Sprechverbin-

dung zwischen mindestens zwei Endgeräten (**2, 3**) übertragen wird, mit den Schritten
Empfangen eines Aktivierungssignals durch eine Steuereinheit (**4**),
Analysieren des ersten Audiosignals, das nach dem Empfang des Aktivierungssignals über die Sprechverbindung übertragen wird, mittels eines Spracherkennungssystems (**5**) und Erkennen eines Wortes aus dem ersten Audiosignal durch das Spracherkennungssystem (**5**),
Übergeben des erkannten Wortes an einen Lexikonserver (**6**) und Übersetzen des erkannten Wortes in eine zweite Sprache durch den Lexikonserver (**6**) und Ausgeben des durch den Lexikonserver (**6**) übersetzten Wortes an einen Nutzer mindestens eines der Endgeräte (**2, 3**).

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet**, dass der Schritt des Ausgebens des durch den Lexikonserver (**6**) übersetzten Wortes das Erzeugen eines zweiten Audiosignals aus dem übersetzten Wort mittels eines Text-zu-Sprache-Wandlers und das Ausgeben des zweiten Audiosignals an den Nutzer umfasst.

3. Verfahren nach Anspruch 2, **dadurch gekennzeichnet**, dass in dem Schritt des Ausgebens des zweiten Audiosignals das zweite Audiosignal in die Sprechverbindung ausgegeben wird.

4. Verfahren nach einem der Ansprüche 1 bis 3, **dadurch gekennzeichnet**, dass in dem Schritt des Ausgebens des durch den Lexikonserver (**6**) übersetzten Wortes das übersetzte Wort zu dem Nutzer übertragen und bei dem Nutzer als Text dargestellt wird, wobei die Übertragung und Ausgabe des übersetzten Wortes vorzugsweise mittels einer Webseite erfolgt.

5. Verfahren nach einem der Ansprüche 1 bis 4, **dadurch gekennzeichnet**, dass das Aktivierungssignal durch eines der Endgeräte (**2, 3**) gesendet wird.

6. Verfahren nach Anspruch 5, **dadurch gekennzeichnet**, dass das Erzeugen eines Aktivierungssignals durch Drücken einer Taste des Endgeräts (**2, 3**) ausgelöst wird.

7. Verfahren nach Anspruch 5 oder 6, **dadurch gekennzeichnet**, dass das Erzeugen eines Aktivierungssignals durch Erkennen eines Schlüsselwortes in dem ersten Audiosignal ausgelöst wird.

8. Verfahren nach Anspruch 7, **dadurch gekennzeichnet**, dass zum Erkennen des Schlüsselwortes das erste Audiosignal durch eine Aktivierungseinheit analysiert und bei Erkennen eines Schlüsselwortes in dem ersten Audiosignal das Aktivierungssignal erzeugt wird.

9. Verfahren nach einem der Ansprüche 5 bis 8, **dadurch gekennzeichnet**, dass das Erzeugen eines Aktivierungssignals durch eine mechanische Bewegung des Endgeräts (2, 3) ausgelöst wird.

10. Verfahren nach einem der Ansprüche 1 bis 9, **dadurch gekennzeichnet**, dass die zweite Sprache basierend auf ein das Erzeugen eines Aktivierungssignals auslösendes Ereignis gewählt wird.

11. Verfahren nach einem der Ansprüche 1 bis 10, **dadurch gekennzeichnet**, dass nach dem Empfangen eines Aktivierungssignals das erste Audiosignal, das von dem das Aktivierungssignal erzeugenden Endgerät (2, 3) ausgesendet wird, aus der Sprechverbindung ausgekoppelt wird.

12. Verfahren nach einem der Ansprüche 3 bis 11, **dadurch gekennzeichnet**, dass während des Schrittes des Ausgebens des zweiten Audiosignals das zweite Audiosignal an jedes der an der Sprechverbindung beteiligten Endgeräte (2, 3) ausgegeben wird.

13. Verfahren nach einem der Ansprüche 3 bis 11, **dadurch gekennzeichnet**, dass bei dem Schritt des Ausgebens des zweiten Audiosignals das zweite Audiosignal nur an das Endgerät (2, 3) ausgegeben wird, durch das das Aktivierungssignal erzeugt wurde.

14. Verfahren nach Anspruch 12 oder 13, **dadurch gekennzeichnet**, dass das zweite Audiosignal mittels eines Übertragungssystems (9) mit dem über die Sprechverbindung übertragenen ersten Audiosignal gemischt wird.

15. Verfahren nach einem der Ansprüche 3 bis 14, **dadurch gekennzeichnet**, dass bei dem Erzeugen des zweiten Audiosignals durch den Text-zu-Sprache-Wandler auf eine Datenbank zurückgegriffen wird, die Informationen zur Stimme eines Nutzers an dem das Aktivierungsereignis auslösenden Endgerätes enthält.

16. Verfahren nach einem der Ansprüche 1 bis 15, **dadurch gekennzeichnet**, dass durch den Lexikonserver (6) für das Übersetzen eines erkannten Wortes auf eine Wörterbuchdatenbank zurückgegriffen wird.

17. Verfahren nach Anspruch 16, **dadurch gekennzeichnet**, dass die Wörterbuchdatenbank basierend auf den in der Sprechverbindung erwarteten Sprachinhalten angepasst wird.

18. Verfahren nach einem der Ansprüche 1 bis 17, **dadurch gekennzeichnet**, dass der Schritt des Analysierens des ersten Audiosignals durch das Spracherkennungssystem (5) bei Empfang des Aktivierungssignals gestartet und durch eine Terminierungssignal beendet wird, wobei das Erzeugen eines

Terminierungssignal vorzugsweise durch das Erkennen einer Wortlücke, durch den Ablauf einer vorgebbaren Zeitspanne, durch Drücken einer Taste an dem das Aktivierungssignal erzeugenden Endgerät, durch das Erkennen eines Schlüsselwortes und/oder durch mechanische Bewegung des das Aktivierungssignal erzeugenden Endgeräts ausgelöst wird.

19. System zum Bereitstellen einer Übersetzung eines Sprachinhalts aus einem ersten Audiosignal, insbesondere zum Durchführen eines Verfahrens nach einem der Ansprüche 1 bis 18, wobei das erste Audiosignal Sprachinhalte in einer ersten Sprache enthält und über eine Sprechverbindung zwischen mindestens zwei Endgeräten (2, 3) übertragen wird, mit

einer Steuereinheit (4), die in die Sprechverbindung einkoppelbar ist und zum Empfang eines Aktivierungssignals ausgestaltet ist,

einem Spracherkennungssystem (5), das kommunizierend mit der Steuereinheit (4) verbunden ist und das zur Analyse des von der Steuereinheit (4) empfangenen ersten Audiosignals und zum Erkennen eines in dem Audiosignal enthaltenen Wortes in einer ersten Sprache ausgestaltet ist,

einem Lexikonserver (6), der kommunizierend mit dem Spracherkennungssystem (5) verbunden ist und der zum Übersetzen des durch das Spracherkennungssystem (5) erkannten Wortes in ein Wort der zweiten Sprache ausgestaltet ist und

einem Ausgabesystem (7), das kommunizierend mit dem Lexikonserver (6) verbunden ist und zum Ausgeben des durch den Lexikonserver (6) übersetzten Wortes an einen Nutzer mindestens eines der Endgeräte (2, 3) ausgestaltet ist.

20. System nach Anspruch 19, **dadurch gekennzeichnet**, dass das Ausgabesystem (7) einen Text-zu-Sprache-Wandler (8) umfasst, der zum Wandeln des von dem Lexikonserver (6) übersetzten Wortes in ein zweites Audiosignal ausgestaltet ist.

21. System nach Anspruch 20, **dadurch gekennzeichnet**, dass das Ausgabesystem (7) zusätzlich ein Übertragungssystem (9) umfasst, das kommunizierend mit dem Text-zu-Sprache-Wandler (8) verbunden ist und das zum Ausgeben des zweiten Audiosignals, vorzugsweise in die Sprechverbindung, ausgestaltet ist.

22. System nach Anspruch 19, **dadurch gekennzeichnet**, dass das Ausgabesystem (8) einen Webserver (10) umfasst, der kommunizierend mit dem Lexikonserver (6) verbunden ist und der eine Webseite für den Nutzer erzeugt und übermittelt, in der das übersetzte Wort enthalten ist.

23. Kommunikationsnetzwerk zum Übertragen von Audiosignalen zwischen mindestens zwei Endgeräten (2, 3) über eine Sprechverbindung, wobei in die

Sprechverbindung ein System nach einem der Ansprüche 19 bis 22 einkoppelbar ist.

Es folgen 2 Seiten Zeichnungen

Anhängende Zeichnungen

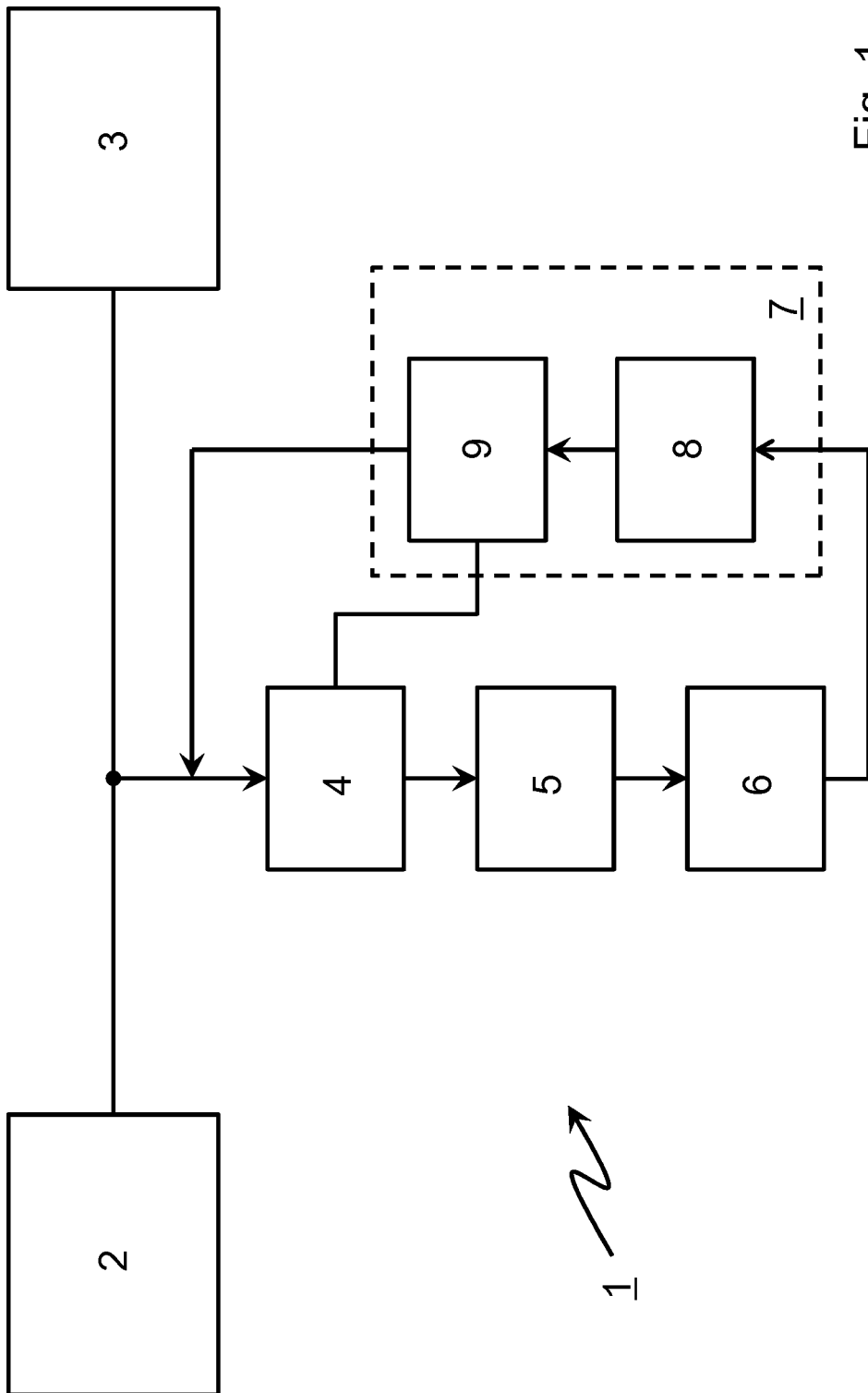


Fig. 1

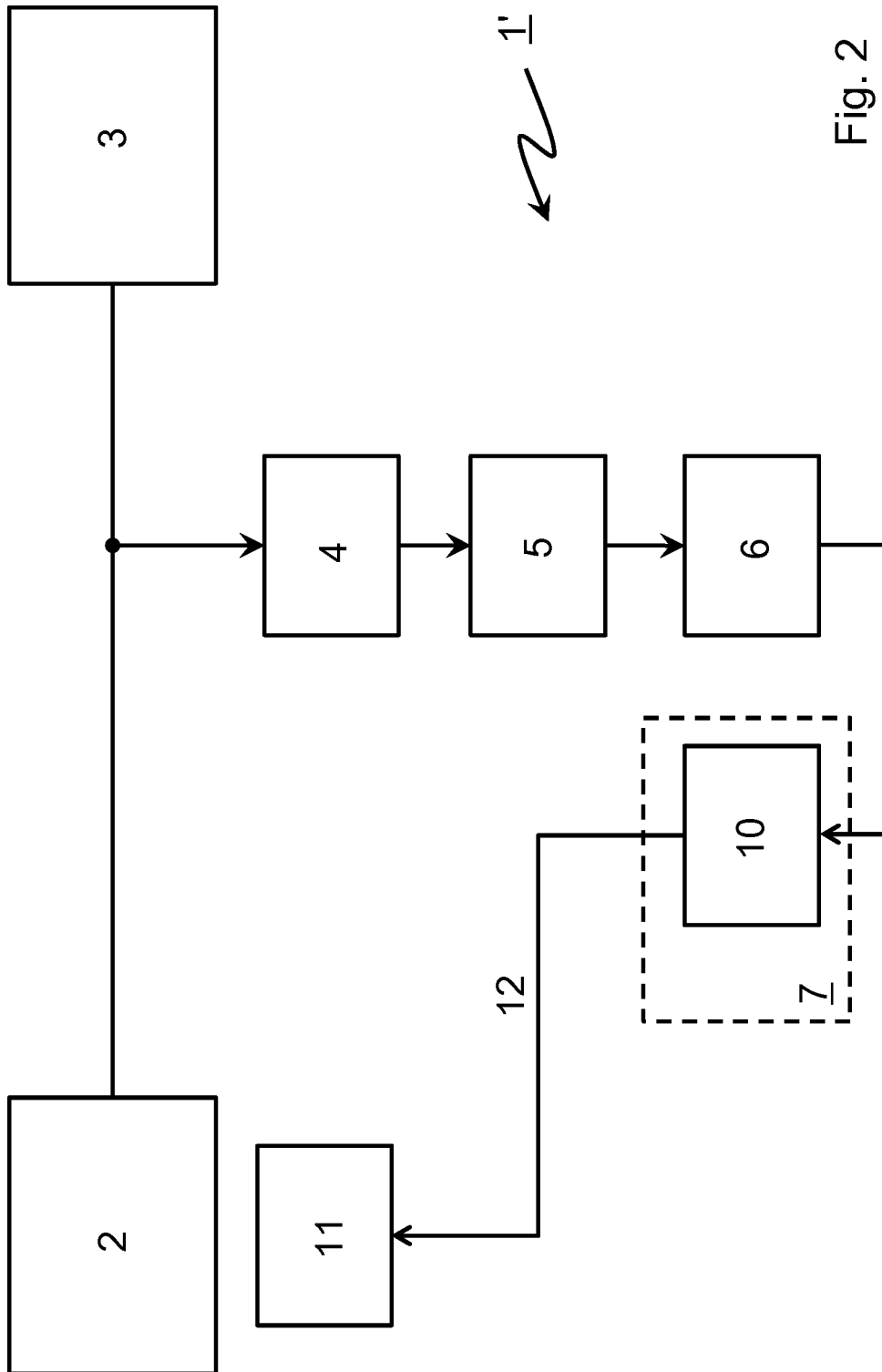


Fig. 2