(19) Deutsches Patent- und Markenamt

(10) **DE 10 2011 120 315 B4** 2013.02.28

(12) # Patentschrift

(21) Aktenzeichen: **10 2011 120 315.3**
(22) Anmeldetag: **06.12.2011**
(43) Offenlegungstag: **14.06.2012**
(45) Veröffentlichungstag der Patenterteilung: **28.02.2013**

(51) Int Cl.: ***G10L 15/06*** (2012.01)

Innerhalb von drei Monaten nach Veröffentlichung der Patenterteilung kann nach § 59 Patentgesetz gegen das Patent Einspruch erhoben werden. Der Einspruch ist schriftlich zu erklären und zu begründen. Innerhalb der Einspruchsfrist ist eine Einspruchsgebühr in Höhe von 200 Euro zu entrichten(§ 6 Patentkostengesetz in Verbindung mit der Anlage zu § 2 Abs. 1 Patentkostengesetz).

(30) Unionspriorität:
12/965,508 **10.12.2010** **US**

(73) Patentinhaber:
General Motors LLC, Detroit, Mich., US

(74) Vertreter:
Manitz, Finsterwald & Partner GbR, 80336, München, DE

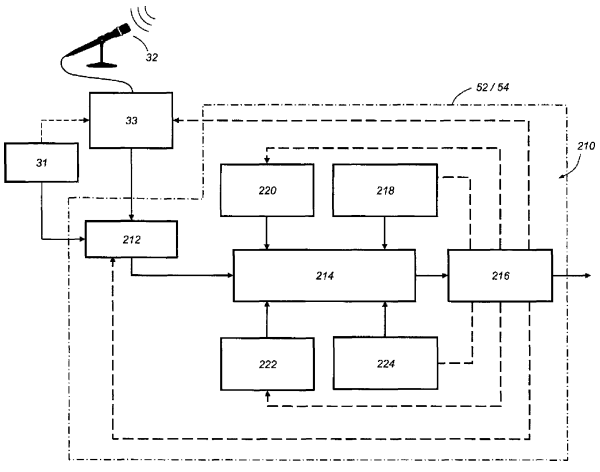
(72) Erfinder:
Talwar, Gaurav, Farmington Hills, Mich., US;
Chengalvarayan, Rathinavelu, Naperville, Ill., US

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

US 2008 / 0 010 057 A1
US 2009 / 0 119 105 A1
US 2009 / 0 198 494 A1
US 2010 / 0 076 765 A1

(54) Bezeichnung: **Anpassung männlicher Akustikmodelle auf der Basis von sprachunabhängigen weiblichen Sprechdaten**

(57) Hauptanspruch: Verfahren zum Erzeugen von Proxy-Akustikmodellen für die Verwendung bei der automatischen Spracherkennung, welches die Schritte umfasst, dass:
(a) ausgehend von Sprachmaterial, das über Mikrophon von männlichen Sprechern einer ersten Sprache gewonnen worden ist, Akustikmodelle trainiert werden, und
(b) die in Schritt (a) trainierten Akustikmodelle in Reaktion auf sprachenunabhängige Sprechdaten von weiblichen Sprechern einer zweiten Sprache angepasst werden, um Proxy-Akustikmodelle zur Verwendung während der Laufzeit der Spracherkennung einer Äußerung eines weiblichen Sprechers der ersten Sprache zu erzeugen.



Beschreibung

TECHNISCHES GEBIET

[0001] Die vorliegende Erfindung betrifft allgemein die automatische Spracherkennung.

HINTERGRUND DER ERFINDUNG

[0002] Automatische Spracherkennungstechnologien (ASE-Technologien) versetzen mit Mikrophon ausgerüstete Datenverarbeitungsgeräte in die Lage, Sprache zu interpretieren, und bieten somit eine Alternative zu herkömmlichen Mensch-Computer-Eingabeeinrichtungen wie beispielsweise Tastaturen oder Tastenfeldern. ASE-Systeme verwenden Akustikmodelle zum Erkennen von Sprache. Ein Akustikmodell ist eine statistische Repräsentation eines oder mehrerer Laute, die eine Sprachäußerung, wie beispielsweise ein Wort, oder ein Phonem oder sonstiges Subwort darstellen. Ein Akustikmodell für eine Äußerung wird durch einen Trainingsprozess erzeugt, der umfasst, dass Tonaufzeichnungen mehrerer Beispiele der Äußerung seitens mehrerer Personen in mehreren Kontexten angefertigt werden und die Äußerungsbeispiele zu einer oder zu mehreren statistischen Repräsentationen der Äußerung zusammengesetzt werden. Beispielsweise können etwa Akustikmodelle für die Ziffern 0–9 durch 50 Männer und 50 Frauen trainiert werden, die jeweils jede Ziffer zehn Mal unter einer oder mehreren Bedingungen aussprechen. Demzufolge stehen anschließend für jede Ziffer 500 weibliche Äußerungsbeispiele und 500 männliche Äußerungsbeispiele zur Verfügung. Sämtliche Äußerungsbeispiele für jede Ziffer können zu einer oder mehreren geschlechtsneutralen statistischen Repräsentationen einer jeden Ziffer zusammengesetzt werden, oder es können die weiblichen Äußerungsbeispiele für jede Ziffer zu einer oder mehreren weiblichen statistischen Repräsentationen einer jeden Ziffer zusammengesetzt werden und es können die männlichen Äußerungsbeispiele für jede Ziffer zu einer oder mehreren männlichen statistischen Repräsentationen einer jeden Ziffer zusammengesetzt werden.

[0003] Ein Problem, das im Zusammenhang mit der ASE auftritt, besteht jedoch darin, dass unter Umständen für die weiblichen Sprecher bestimmter Bevölkerungsgruppen nur wenig bis überhaupt keine Trainingsdaten verfügbar sind. In bestimmten Bevölkerungsgruppen kann es beispielsweise schwierig oder sogar unmöglich sein, Trainingsdaten für weibliche Akustikmodelle zu erhalten. Als weiteres Beispiel nehmen in gewissen Bevölkerungsgruppen viele weibliche Personen derzeit nicht als Autofahrer am Verkehrsgeschehen teil, so dass es an statistisch signifikanten weiblichen Sprechdaten in Fahrzeuginnenräumen mangelt.

[0004] Aus der US 2009/0119105 A1 ist ein Verfahren zum Erzeugen von Proxy-Akustikmodellen für die Verwendung bei der automatischen Spracherkennung bekannt, bei dem ausgehend von Sprachmaterial, das über Mikrophon von Sprechern einer ersten Sprache gewonnen worden ist, Akustikmodelle trainiert werden und die trainierten Akustikmodelle in Reaktion auf Sprechdaten von Sprechern einer zweiten Sprache angepasst werden, um Proxy-Akustikmodelle zur Verwendung während der Laufzeit der Spracherkennung zu erzeugen.

[0005] Ein Verfahren mit entsprechenden Schritten ist in der US 2010/0076765 A1 beschrieben.

[0006] Die US 2008/0010057 A1 und die US 2009/0198494 A1 beschreiben weitere Verfahren zur automatischen Anpassung eines Spracherkennungssystems, bei denen Akustikmodelle ausgehend von Sprachmaterial trainiert bzw. Parameter zur Anpassung der Spracherkennung ermittelt werden.

[0007] Es ist eine Aufgabe der Erfindung, ein Verfahren zu schaffen, um die Leistungsfähigkeit der Spracherkennung für bestimmte weibliche Benutzer zu verbessern, auch wenn ein Mangel an weiblichen Sprechdaten für das Training der Spracherkennung vorliegt.

KURZFASSUNG DER ERFINDUNG

[0008] Diese Aufgabe wird gelöst durch ein Verfahren mit den Merkmalen der Ansprüche 1 oder 9.

[0009] Gemäß einer Ausführungsform der Erfindung wird ein Verfahren zum Erzeugen von Proxy-Akustikmodellen für die Verwendung bei der automatischen Spracherkennung bereitgestellt. Das Verfahren umfasst die Schritte, dass (a) ausgehend von Sprachmaterial, das über Mikrophon von männlichen Sprechern einer ersten Sprache gewonnen worden ist, Akustikmodelle trainiert werden, und (b) die in Schritt (a) trainierten Akustikmodelle in Reaktion auf sprachenunabhängige Sprechdaten von weiblichen Sprechern einer zweiten Sprache angepasst werden, um Proxy-Akustikmodelle zur Verwendung während der Laufzeit der Spracherkennung einer Äußerung eines weiblichen Sprechers der ersten Sprache zu erzeugen.

[0010] Gemäß einer anderen Ausführungsform der Erfindung wird ein Verfahren zur automatischen Spracherkennung bereitgestellt, welches die Schritte umfasst, dass: (a) eine Äußerung über ein Mikrophon von einem weiblichen Sprecher einer ersten Sprache empfangen wird; (b) die Äußerung mit einem Präprozessor zur automatischen Spracherkennung vorverarbeitet wird, um Akustikmerkmalsvektoren zu erzeugen; (c) zumindest eine Formantenfrequenz der empfangenen Äußerung bestimmt wird; (d) zumindest ei-

nes aus einer Mehrzahl von Formantenfrequenzbändern in Sprechdaten von weiblichen Sprechern einer zweiten Sprache identifiziert wird, das der zumindest einen in Schritt (c) bestimmten Formantenfrequenz entspricht; und (e) Akustikmodelle, die ausgehend von Sprachmaterial von männlichen Sprechern der ersten Sprache trainiert worden sind, in Reaktion auf den Identifikationsschritt (d) angepasst werden, um als Ergebnis Proxy-Akustikmodelle für die weiblichen Sprecher der ersten Sprache zu liefern.

KURZBESCHREIBUNG DER ZEICHNUNGEN

[0011] Eine oder mehrere bevorzugte, beispielhafte Ausführungsformen der Erfindung werden im Folgenden in Verbindung mit den beigefügten Zeichnungen beschrieben, in denen gleiche Bezeichnungen jeweils gleiche Elemente bezeichnen, und wobei:

[0012] Fig. 1 ein Blockdiagramm ist, in welchem eine beispielhafte Ausführungsform eines Kommunikationssystems abgebildet ist, mit dem das hier offenbarte Verfahren verwendet werden kann;

[0013] Fig. 2 ein Blockdiagramm ist, in welchem eine beispielhafte Ausführungsform eines automatischen Spracherkennungssystems (ASE-Systems) veranschaulicht ist, das mit dem System aus Fig. 1 zum Einsatz kommen kann und zur Implementierung beispielhafter Verfahren der Spracherkennung verwendet wird;

[0014] Fig. 3 ein Ablaufdiagramm ist, in welchem eine beispielhafte Ausführungsform eines Verfahrens zum Erzeugen von Akustikmodellen für die Verwendung bei der durch das ASE-System aus Fig. 2 ausführbaren, automatischen Spracherkennung veranschaulicht ist; und

[0015] Fig. 4 ein Ablaufdiagramm ist, in welchem eine beispielhafte Ausführungsform eines Verfahrens zur automatischen Spracherkennung veranschaulicht ist, das durch das ASE-System aus Fig. 2 ausgeführt werden kann.

DETAILLIERTE BESCHREIBUNG DER BEVORZUGTEN AUSFÜHRUNGSFORM(EN)

[0016] Die nachfolgende Beschreibung beschreibt ein beispielhaftes Kommunikationssystem, ein beispielhaftes ASE-System, das mit dem Kommunikationssystem verwendet werden kann, sowie ein oder mehrere beispielhafte Verfahren, die mit einem oder mit beiden der vorerwähnten Systeme verwendet werden können. Die weiter unten beschriebenen Verfahren können von einer Fahrzeugtelematikeinheit (FTE) als Teil der Erkennung von Sprachäußerungen eines Benutzers der FTE verwendet werden. Die weiter unten beschriebenen Verfahren sind zwar so dargestellt, wie sie in einer FTE implementiert werden könnten, es ist jedoch festzustellen, dass sie in jeder beliebigen Art von Fahrzeug-Spracherkennungssystem, sowie auch in anderen Arten von Spracherkennungssystemen zweckmäßig sein könnten. Beispielsweise könnten die Verfahren etwa in ASE-fähigen mobilen Computereinrichtungen oder -systemen, in PCs oder dergleichen implementiert sein.

Kommunikationssystem

[0017] In Fig. 1, auf welche nun Bezug genommen wird, ist eine beispielhafte Betriebsumgebung gezeigt, welche ein mobiles Fahrzeugkommunikationssystem **10** umfasst und zur Implementierung des hier offenbarten Verfahrens verwendet werden kann. Das Kommunikationssystem **10** umfasst allgemein ein Fahrzeug **12**, ein oder mehrere Mobilfunknetzsysteme **14**, ein terrestrisches Kommunikationsnetz **16**, einen Computer **18** und ein Callcenter **20**. Es versteht sich, dass das offenbarte Verfahren mit jeder beliebigen Anzahl von verschiedenen Systemen verwendet werden kann und nicht spezifisch auf die hier gezeigte Betriebsumgebung beschränkt ist. Außerdem sind die Architektur, der Aufbau, die Konfiguration und der Betrieb des Systems **10** und seiner einzelnen Bauteile gemäß dem Stand der Technik bekannt. Somit bieten die folgenden Absätze lediglich einen kurzen Überblick über ein solches beispielhaftes System **10**; andere, hier nicht gezeigte Systeme könnten jedoch ebenso das offenbarte Verfahren verwenden.

[0018] Das Fahrzeug **12** ist in der veranschaulichten Ausführungsform als ein Personenkraftwagen abgebildet, es ist jedoch festzustellen, dass jedes beliebige andere Fahrzeug, einschließlich Motorräder, Lastkraftwagen, Geländewagen (SUVs), Wohnmobile (RVs), Wasserfahrzeuge, Luftfahrzeuge, usw. ebenso verwendet werden können. Ein Teil der Fahrzeugelektronik **28** ist allgemein in Fig. 1 gezeigt und umfasst eine Telematikeinheit **30**, ein Mikrophon **32**, eine oder mehrere Tasten oder andere Steuereingänge **34**, ein Audiosystem **36**, ein Anzeigedisplay **38** und ein GPS-Modul **40** sowie eine Anzahl von Fahrzeugsystemmodulen (FSMs) **42**. Manche dieser Einrichtungen können direkt mit der Telematikeinheit verbunden sein, wie beispielsweise das Mikrophon **32** und die Drucktaste(n) **34**, während andere indirekt unter Verwendung einer oder mehrerer Netzverbindungen, wie beispielsweise eines Kommunikationsbusses **44** oder eines Unterhaltungsbusses **46**, verbunden sind. Als Beispiele für geeignete Netzwerkverbindungen sind unter anderem zu nennen: ein Controller Area Network (CAN), ein Media Oriented System Transport (MOST), ein Local Interconnect Network (LIN), ein Local Area Network (LAN), sowie andere geeignete Verbindungen, wie beispielsweise Ethernet oder andere, welche den bekannten ISO-, SAE- und IEEE-Normen und Spezifikationen entsprechen, um hier nur einige zu nennen.

[0019] Bei der Telematikeinheit **30** kann es sich um eine vom Erstausrüster installierte (eingebettete) OEM-Einrichtung oder um eine nachträglich eingebaute Einrichtung handeln, die eine drahtlose Sprach- und/oder Datenkommunikation über das Mobilfunknetzsystem **14** und über Funknetzwerkverbindungen ermöglicht, so dass das Fahrzeug mit dem Callcenter **20** oder einer anderen Einheit oder Einrichtung kommunizieren kann. Die Telematikeinheit verwendet zur Herstellung eines Kommunikationskanals (eines Sprachkanals und/oder Datenkanals) mit dem Mobilfunknetzsystem **14** vorzugsweise Funkübertragungen, so dass Sprach- und/oder Datenübertragungen über den Kanal gesendet oder empfangen werden können. Indem sie gleichermaßen Sprach- und Datenkommunikation bereitstellt, ermöglicht es die Telematikeinheit **30**, dass in dem Fahrzeug eine Reihe verschiedener Dienste, einschließlich jener im Bereich der Navigation, der Telefonie, der Notfallunterstützung, der Diagnose, der informierenden Unterhaltung (Infotainment), usw. angeboten werden können. Daten können entweder über eine Datenverbindung, beispielsweise als Paketdatenübertragung über einen Datenkanal, oder über einen Sprachkanal unter Verwendung von in der Technik bekannten Verfahren gesendet werden. Für kombinierte Dienste, bei denen gleichermaßen Sprachkommunikation (z. B. mit einem menschlichen Berater oder einer VoiceResponse-Einheit in einem Callcenter **20**) und Datenkommunikation (z. B. zur Bereitstellung von GPS-Positionsdaten oder Fahrzeugdiagnosedaten an das Callcenter **20**) benötigt wird, kann das System einen einzelnen Anruf über einen Sprachkanal verwenden und je nach Bedarf zwischen Sprach- und Datenübertragung über den Sprachkanal wechseln, und dies kann unter Verwendung von Verfahren erfolgen, die dem Fachmann bekannt sind.

[0020] Gemäß einer Ausführungsform verwendet die Telematikeinheit **30** zellulare Kommunikation gemäß entweder GSM- oder CDMA-Standards und umfasst daher einen zellularen Standard-Chipsatz **50** für Sprachkommunikationsverbindungen wie Telefonieren mit Freisprechanlage, ein Funk-Modem für die Datenübertragung, eine elektronische Verarbeitungseinrichtung **52**, eine oder mehrere digitale Speichereinrichtungen **54** und eine Dualantenne **56**. Es ist festzustellen, dass das Modem entweder als in der Telematikeinheit gespeicherte Software implementiert ist und von dem Prozessor **52** ausgeführt wird, oder als separate Hardwarekomponente vorhanden sein kann, welche innerhalb oder außerhalb der Telematikeinheit **30** angeordnet ist. Das Modem kann unter Verwendung jeder beliebigen Anzahl von unterschiedlichen Standards oder Protokollen, wie beispielsweise EVDO, CDMA, GPRS und EDGE betrieben werden. Funknetzwerkverbindungen zwischen dem Fahrzeug und anderen vernetzten Einrichtungen können ebenfalls unter Verwendung der Telematikeinheit **30** realisiert werden. Zu diesem Zweck kann die Telematikeinheit **30** so ausgelegt sein, dass sie drahtlos gemäß einem oder mehreren Funkprotokollen, wie beispielsweise einem beliebigen der IEEE 802.11-Protokolle, WiMAX oder Bluetooth kommuniziert. Bei einer Verwendung für paketvermittelte Datenkommunikation, wie beispielsweise TCP/IP, kann die Telematikeinheit mit einer statischen IP-Adresse ausgelegt sein oder so eingerichtet sein, dass sie eine von einer anderen Einrichtung im Netzwerk, wie beispielsweise einem Router oder einem Netzwerkadressserver, eine automatisch zugewiesene IP-Adresse erhält.

[0021] Bei dem Prozessor **52** kann es sich um jede beliebige Art von Einrichtung handeln, die in der Lage ist, elektronische Anweisungen zu verarbeiten, also beispielsweise um Mikroprozessoren, Mikrocontroller, Hostprozessoren, Controller, Fahrzeugkommunikationsprozessoren und anwendungsspezifische Schaltkreise (ASICs). Es kann dies ein fest zugeordneter Prozessor sein, der ausschließlich für die Telematikeinheit **30** verwendet wird, oder er kann auch von anderen Fahrzeugsystemen gemeinsam genutzt werden. Der Prozessor **52** führt verschiedene Arten von digital gespeicherten Anweisungen aus, beispielsweise in dem Speicher **54** abgespeicherte Software- oder Firmwareprogramme, welche die Telematikeinheit in die Lage versetzen, ein breites Spektrum von Diensten bereitzustellen. Zum Beispiel kann der Prozessor **52** Programme oder Prozessdaten ausführen, um zumindest einen Teil des hier besprochenen Verfahrens durchzuführen.

[0022] Die Telematikeinheit **30** kann dazu verwendet werden, unterschiedlichste Fahrzeug-Dienste bereitzustellen, die eine Funkkommunikation zu und/oder von dem Fahrzeug erforderlich machen. Solche Dienste umfassen: Abbiegehinweise und andere navigationsbezogene Dienste, die in Verbindung mit dem GPS-basierten Fahrzeugnavigationsmodul **40** bereitgestellt werden; Airbag-Auslösungsbenachrichtigung und andere notfallunterstützungs- oder pannenhilfebezogene Dienste, die in Verbindung mit einem oder mehreren Kollisionssensorschnittstellenmodulen wie beispielsweise einem Karosserie-Steuermodul (nicht gezeigt) bereitgestellt werden; Diagnoseberichtswesen mit einem oder mehreren Diagnosemodulen; sowie infotainmentbezogene Dienste, wobei Musik, Webseiten, Filme, Fernsehprogramme, Videospiele und/oder andere Daten durch ein Infotainment-Modul (nicht gezeigt) heruntergeladen und im Hinblick auf eine direkte oder spätere Wiedergabe abgespeichert werden. Bei den weiter oben aufgelisteten Diensten handelt es sich keinesfalls um eine erschöpfende Liste aller Fähigkeiten der Telematikeinheit **30** sondern lediglich um eine Aufzählung einiger der Dienste, welche die Telematikeinheit anzubieten in der Lage ist. Weiterhin versteht es sich, dass zumindest manche der weiter oben erwähnten

Module in Form von Softwarebefehlen implementiert sein könnten, die innerhalb oder außerhalb der Telematikeinheit **30** abgespeichert sind, oder als Hardwarekomponenten vorhanden sein könnten, die innerhalb oder außerhalb der Telematikeinheit **30** untergebracht sein können, oder sie könnten miteinander oder mit anderen Systemen, die über das gesamte Fahrzeug hinweg verteilt sein können, integriert oder gemeinsam genutzt sein, um nur einige Möglichkeiten zu nennen. Falls die Module als außerhalb der Telematikeinheit **30** untergebrachte Fahrzeugsystemmodule (FSMs) **42** implementiert sind, könnten sie den Fahrzeugbus **44** verwenden, um Daten und Anweisungen mit der Telematikeinheit auszutauschen.

[0023] Das GPS-Modul **40** empfängt Funksignale von einer Konstellation **60** aus GPS-Satelliten. Aus diesen Signalen kann das Modul **40** die Fahrzeugposition bestimmen, welche zur Bereitstellung von navigations- und sonstigen positionsbezogenen Diensten an den Fahrzeuglenker verwendet wird. Navigationsbezogene Daten können auf dem Display **38** (oder einem anderen in dem Fahrzeug vorhandenen Display) angezeigt werden oder können verbal erfolgen, wie dies etwa im Fall der Bereitstellung einer auch als Turn-by-Turn-Navigation bezeichneten, sprachgesteuerten Routenführung geschieht. Diese Navigationsdienste können unter Verwendung eines fest zugeordneten, fahrzeugeigenen Navigationsmoduls (welches Teil des GPS-Moduls **40** sein kann) bereitgestellt werden, oder es können manche oder alle navigationsbezogenen Dienste über die Telematikeinheit **30** erfolgen, wobei die Positionsdaten zum Zweck der Versorgung des Fahrzeugs mit Navigationskarten, Kartenanmerkungen (als 'Points of Interest' bezeichnete Sonderziele, Restaurants, usw.), Routenberechnungen und dergleichen an eine entfernte Stelle gesendet werden. Die Positionsdaten können auch für andere Zwecke, wie etwa die Fuhrparkverwaltung, an das Callcenter **20** oder an ein entferntes Computersystem, wie beispielsweise den Computer **18**, bereitgestellt werden. Außerdem können neue oder aktualisierte Kartendaten für das GPS-Modul **40** über die Telematikeinheit **30** von dem Callcenter **20** heruntergeladen werden.

[0024] Abgesehen von dem Audiosystem **36** und dem GPS-Modul **40** kann das Fahrzeug **12** weitere Fahrzeugsystemmodule (FSMs) **42** in Form von elektronischen Hardwarekomponenten umfassen, die über das gesamte Fahrzeug hinweg angeordnet sind und in der Regel Eingangsdaten von einem oder mehreren Sensoren empfangen und die empfangenen Eingangsdaten dazu verwenden, um Diagnose-, Überwachungs-, Steuer-, Berichts- und/oder sonstige Funktionen auszuführen. Ein jedes der FSMs **42** ist vorzugsweise über den Kommunikationsbus **44** mit den anderen FSMs sowie mit der Telematikeinheit **30** verbunden und kann so programmiert sein, dass es Diagnosetests von Fahrzeugsystemen und -subsystemen durchführt. Beispielsweise kann ein FSM **42** ein Motorsteuermodul (ECM) darstellen, das verschiedene Aspekte des Motorbetriebs, wie etwa die Kraftstoffzündung und den Zündzeitpunkt steuert, ein anderes FSM **42** kann ein Antriebsstrang-Steuermodul darstellen, das den Betrieb einer oder mehrerer Komponenten des Fahrzeugantriebsstrangs steuert, und ein weiteres FSM **42** kann ein Karosserie-Steuermodul darstellen, das verschiedene über das gesamte Fahrzeug hinweg verteilte elektrische Komponenten, wie etwa die Zentralverriegelung und die Scheinwerfer des Fahrzeugs, regelt. Gemäß einer Ausführungsform ist das Motorsteuermodul mit On-Board-Diagnosemerkmalen (OBD-Merkmalen) ausgestattet, die eine große Anzahl von Echtzeitdaten bereitstellen, wie beispielsweise jene, die von verschiedenen Sensoren einschließlich der Fahrzeugabgassensoren, empfangen werden und eine standardisierte Reihe von Störfall-Diagnosecodes (Diagnostic Trouble Codes, DTCs) bereitstellen, welche es einem Techniker erlauben, Funktionsstörungen innerhalb des Fahrzeugs rasch aufzufinden und zu beheben. Wie für den Fachmann festzustellen ist, handelt es sich bei den oben erwähnten FSMs nur um Beispiele einiger Module, die in dem Fahrzeug **12** verwendet werden können, wobei zahlreiche weitere Module ebenso möglich sind.

[0025] Die Fahrzeugelektronik **28** umfasst außerdem eine Anzahl von Fahrzeugbenutzerschnittstellen, durch die dem Fahrzeugbenutzer Mittel zum Bereitstellen und/oder Empfangen von Informationen, einschließlich des Mikrophons **32**, der Drucktaste(n) **34**, des Audiosystems **36** und des Anzeigedisplays **38**, zur Verfügung gestellt werden. Der hier verwendete Begriff 'Fahrzeugbenutzerschnittstelle' umfasst ganz allgemein jede geeignete Form von elektronischer Einrichtung, einschließlich Hardwarekomponenten und Softwarekomponenten, die sich an Bord des Fahrzeugs befindet und es einem Fahrzeugbenutzer ermöglicht, mit oder über eine Komponente des Fahrzeugs zu kommunizieren. Das Mikrophon **32** stellt einen Audioeingang zu der Telematikeinheit bereit, um das Eingeben von Sprachbefehlen sowie das Freisprechen über das Mobilfunknetzsystem **14** durch den Fahrer oder den Beifahrer zu ermöglichen. Zu diesem Zweck kann es mit einer eingebauten, automatisierten Sprachverarbeitungseinheit verbunden sein, welche eine in der Technik bekannte Mensch/Maschine-Schnittstellentechnologie ('Human-Machine Interface'- oder HMI-Technologie) verwendet. Durch die Drucktaste(n) **34** wird eine manuelle Benutzereingabe an die Telematikeinheit **30** ermöglicht, um Funktelefonanrufe einzuleiten und andere Daten-, Rückmelde- oder Steuereingaben vorzunehmen. Es können separate Drucktasten verwendet werden, um Notfallanrufe gegenüber regulären Kundendienstanrufen an das Callcenter **20** abzuheben. Das Audiosystem **36** stellt einen Audioausgang

für einen Fahrzeuginsassen bereit und es kann sich dabei um ein fest zugeordnetes, unabhängiges System oder um einen Teil des primären Fahrzeugaudiosystems handeln. Gemäß der hier gezeigten, besonderen Ausführungsform ist das Audiosystem **36** funktional gleichermaßen mit dem Fahrzeugbus **44** und dem Unterhaltungsbus **46** gekoppelt und somit in der Lage, MW-, UKW- und Satellitenradio, sowie CD-, DVD- und andere Multimedia-Funktionalität bereitzustellen. Diese Funktionalität kann in Verbindung mit dem weiter oben beschriebenen Infotainment-Modul oder unabhängig von diesem bereitgestellt werden. Das Anzeigedisplay **38** ist vorzugsweise als ein Grafikdisplay, wie etwa ein Touchscreen auf der Instrumententafel, oder als eine von der Windschutzscheibe reflektierte, auch als Heads-up Display bezeichnete Überkopfanzeige ausgebildet und kann dazu benutzt werden, eine Vielzahl von Eingabe- und Ausgabefunktionen bereitzustellen. Verschiedene andere Fahrzeugbenutzerschnittstellen können ebenfalls zum Einsatz kommen, da die Schnittstellen aus Fig. 1 nur ein Beispiel für eine besondere Realisierung darstellen.

[0026] Bei dem Mobilfunknetzsystem **14** handelt es sich vorzugsweise um ein Zellulartelefonsystem, das eine Mehrzahl von Mobilfunkmasten **70** (nur einer gezeigt), ein oder mehrere Mobilvermittlungsstellen (Mobile Switching Centers, MSCs) **72** sowie jedwede anderen Netzwerkkomponenten umfasst, die erforderlich sind, um das Mobilfunknetzsystem **14** mit dem terrestrischen Netz **16** zu verbinden. Jeder Mobilfunkmast **70** umfasst Sende- und Empfangsantennen und eine Basisstation, wobei die Basisstationen von verschiedenen Mobilfunkmasten jeweils entweder direkt oder über eine zwischengeschaltete Ausrüstung, wie beispielsweise einen Basisstationscontroller, mit der Mobilvermittlungsstelle (MSC) **72** verbunden sind. In dem Zellularsystem **14** kann jede beliebige geeignete Kommunikationstechnologie implementiert sein, beispielsweise analoge Technologien, wie etwa AMPS, oder die neueren digitalen Technologien, wie beispielsweise CDMA (z. B. CDMA2000) oder GMS/GPRS. Wie für den Fachmann festzustellen ist, sind verschiedene Anordnungen bestehend aus Mobilfunkmast, Basisstation und Mobilvermittlungsstelle (MSC) möglich, die jeweils mit dem Funknetzsystem **14** verbunden werden könnten. Beispielsweise könnten sich die Basisstation und der Mobilfunkmast gemeinsam am selben Standort befinden, oder sie könnten voneinander entfernt gelegen sein, jede Basisstation könnte für einen einzelnen Mobilfunkmast verantwortlich sein, oder ein und dieselbe Basisstation könnte verschiedene Mobilfunkmasten bedienen, und verschiedene Basisstationen könnten mit ein und derselben Mobilvermittlungsstelle (MSC) gekoppelt sein, um nur ein paar der möglichen Anordnungen zu nennen.

[0027] Neben der Verwendung des Mobilfunknetzsystems **14** kann ein anderes Mobilfunknetzsystem in Form einer satellitengestützten Kommunikation verwendet werden, um eine unidirektionale oder bidirektionale Kommunikation mit dem Fahrzeug bereitzustellen. Dies kann unter Verwendung eines oder mehrerer Kommunikationssatelliten **62** und einer Uplink-Sendestation **64** erfolgen. Bei unidirektionaler Kommunikation kann es sich beispielsweise um Satellitenradiodienste handeln, wobei der Programminhalt (Nachrichten, Musik, usw.) von der Sendestation **64** empfangen, paketweise für den Upload aufbereitet und anschließend an den Satelliten **62** gesendet wird, der das Programm an die Abonnenten übermittelt. Bei bidirektionaler Kommunikation kann es sich beispielsweise um Satellitentelefoniedienste handeln, die den Satelliten **62** dazu verwenden, um telefonische Mitteilungen zwischen dem Fahrzeug **12** und der Station **64** zu übertragen. Diese Satellitentelefonie kann, sofern sie benutzt wird, entweder zusätzlich zu dem Mobilfunknetzsystem **14** oder anstelle von diesem verwendet werden.

[0028] Bei dem terrestrischen Netz **16** kann es sich entweder um ein herkömmliches landgestütztes Telekommunikationsnetz handeln, das mit einem oder mehreren Festnetztelefonen verbunden ist und das Mobilfunknetzsystem **14** mit dem Callcenter **20** verbindet. Beispielsweise kann das terrestrische Netz **16** ein öffentliches Fernsprechvermittlungsnetz (public switched telephone network, PSTN) umfassen, wie etwa jenes, das zur Bereitstellung der Festnetztelefonie, der paketvermittelten Datenkommunikation und der Internet-Infrastruktur verwendet wird. Eines oder mehrere Segmente des terrestrischen Netzes **16** könnten durch die Verwendung eines standardmäßigen drahtgebundenen Netzes, eines Fasernetzes oder eines anderen optischen Netzes, eines Kabelnetzes, eines Stromleitungsnetzes, anderer drahtloser Netze wie beispielsweise drahtloser lokaler Netze (Wireless Local Area Networks, WLANs) oder von Netzen, die einen drahtlosen Breitbandzugang (Broadband Wireless Access, BWA) ermöglichen, oder jeder beliebigen Kombination aus diesen, implementiert sein. Überdies braucht das Callcenter **20** nicht über ein terrestrisches Netz **16** verbunden sein, sondern es könnte auch eine drahtlose Telefonieausrüstung umfassen, so dass es unmittelbar mit einem Funknetz wie beispielsweise dem Mobilfunknetzsystem **14** kommunizieren kann.

[0029] Bei dem Computer **18** kann es sich um einen aus einer Anzahl von Computern handeln, die über ein privates oder öffentliches Netz, wie beispielsweise das Internet, zugänglich ist. Ein Computer **18** kann für einen oder mehrere Zwecke eingesetzt werden, beispielsweise als Webserver, der von dem Fahrzeug aus über die Telematikeinheit **30** und das Funknetz **14** zugänglich ist. Bei anderen dergestalt zugänglichen Computer **18** kann es sich beispielswei-

se um die folgenden handeln: einen Servicecenter-Computer, zu welchem Diagnosedaten und andere Fahrzeugdaten von dem Fahrzeug über die Telematikeinheit **30** hochgeladen werden können; einen Clientcomputer, der von dem Fahrzeugbesitzer oder einem anderen Abonnenten zu Zwecken wie etwa dem Zugriff auf Fahrzeugdaten oder dem Empfangen von solchen oder zum Einrichten oder Konfigurieren von bevorzugten Abonnenteneinstellungen oder zum Steuern von Fahrzeugfunktionen verwendet wird; oder ein Repository eines Drittanbieters, von welchem oder an welches durch Kommunikationsaufbau mit dem Fahrzeug **12** oder dem Callcenter **20**, oder mit beiden, Fahrzeugdaten oder andere Informationen bereitgestellt werden. Ein Computer **18** kann auch dazu benutzt werden, eine Internetanbindung bereitzustellen, beispielsweise in Form von DNS-Diensten oder eines Netzwerkadressservers, der DHCP oder ein anderes geeignetes Protokoll verwendet, um dem Fahrzeug **12** eine IP-Adresse zuzuweisen.

[0030] Das Callcenter **20** ist so konzipiert, dass es der Fahrzeugelektronik **28** eine Reihe unterschiedlicher, systemspezifischer Backend-Funktionen bereitstellt, und umfasst allgemein, gemäß der hier gezeigten beispielhaften Ausführungsform, einen oder mehrere Schalter **80**, Server **82**, Datenbanken **84**, menschliche Berater **86**, sowie auch ein automatisiertes Voice-Response-System **88**, welche alle in der Technik bekannt sind. Diese verschiedenen Callcenter-Komponenten sind vorzugsweise über ein drahtgebundenes oder drahtloses lokales Netz **90** miteinander gekoppelt. Der Schalter **80**, bei dem es sich um einen Nebenanschlussschalter (Private Branch Exchange- oder PBX-Schalter) handeln kann, leitet die eingehenden Signale so, dass Sprachübertragungen entweder über ein gewöhnliches Telefon an den menschlichen Ansprechpartner **86** oder mittels VoIP an das automatisierte VoiceResponse-System **88** gesendet werden. Das Telefon des menschlichen Beraters kann ebenfalls VoIP verwenden, wie dies durch die gestrichelte Linie in Fig. 1 angedeutet ist. VoIP oder jede andere Datenübertragung durch den Schalter **80** ist über ein zwischen dem Schalter **80** und dem Netz **90** geschaltetes Modem (nicht gezeigt) implementiert. Die Datenübertragungen werden über das Modem zu dem Server **82** und/oder der Datenbank **84** geleitet. In der Datenbank **84** können Kontoinformationen wie beispielsweise Abonnentenauthentifizierungsdaten, Fahrzeugkennungen, Profildatensätze, Verhaltensmuster und andere relevante Abonnentendaten abgespeichert sein. Datenübertragungen können auch durch Funknetzsysteme, wie beispielsweise 802.11x, GPRS und dergleichen durchgeführt werden. Die veranschaulichte Ausführungsform wurde zwar so beschrieben, als würde sie in Verbindung mit einem personalbesetzten Callcenter **20** und unter Mitwirkung eines menschlichen Beraters **86** verwendet, es ist jedoch festzustellen, dass in dem Callcenter statt dessen auch das VRS **88** als ein automatisierter Berater verwendet werden kann oder dass eine Kombination aus VRS **88** und dem menschlichen Berater **86** zum Einsatz kommen kann.

System zur automatischen Spracherkennung

[0031] In Fig. 2, auf welche nun Bezug genommen wird, ist eine beispielhafte Architektur für ein ASE-System **210** gezeigt, das zur Umsetzung des hier offenbarten Verfahrens verwendet werden kann. Ein Fahrzeuginsasse tritt im Allgemeinen zu einem oder mehreren der folgenden grundlegenden Zwecke mit einem System zur automatischen Spracherkennung (ASE-System) über die Sprache in Interaktion: zum Trainieren des Systems, damit es die Besonderheiten der Stimme eines Fahrzeuginsassen erkennt; zum Speichern von diskretem Sprachmaterial, beispielsweise eines gesprochenen Namensetiketts oder eines gesprochenen Steuerworts, wie etwa einer Zahl oder eines Schlüsselworts; oder zum Erkennen der Sprache des Fahrzeuginsassen im Hinblick auf jeden beliebigen geeigneten Zweck, wie beispielsweise sprachgesteuertes Wählen, Menünavigation, Transkription, Service-Anforderungen, Fahrzeugeinrichtungssteuerung oder Einrichtungsfunktionssteuerung oder dergleichen. Im Allgemeinen extrahiert die ASE Akustikdaten aus der menschlichen Sprache, vergleicht die Akustikdaten mit gespeicherten Subwortdaten und stellt sie diesen gegenüber, wählt ein geeignetes Subwort aus, das mit anderen ausgewählten Subwörtern verknüpft werden kann, und gibt die verknüpften Subwörter oder Wörter zur Nachverarbeitung, wie beispielsweise zum Diktat oder zur Transkription, zum adressbuchgesteuerten Wählen, zum Abspeichern im Speicher, zum Trainieren von ASE-Modellen oder Abstimmungsparametern oder dergleichen, aus.

[0032] ASE-Systeme sind dem Fachmann im Allgemeinen bekannt und in Fig. 2 ist nur ein spezifisches, beispielhaftes ASE-System **210** veranschaulicht. Das System **210** umfasst eine Einrichtung zum Empfangen von Sprache, wie beispielsweise das Telematikmikrophon **32**, und eine Akustikschnittstelle **33**, wie beispielsweise eine Soundkarte der Telematikeinheit **30** mit einem Analog-Digital-Wandler, um den Sprachfluss in Akustikdaten zu digitalisieren. Das System **210** umfasst außerdem einen Speicher, wie beispielsweise den Telematikspeicher **54**, zum Speichern der Akustikdaten und zum Speichern von Software und Datenbanken zur Spracherkennung, sowie einen Prozessor, wie beispielsweise den Telematikprozessor **52**, zum Verarbeiten der Akustikdaten. Der Prozessor funktioniert zusammen mit dem Speicher und in Verbindung mit den folgenden Modulen: einem oder mehreren Frontend-Prozessoren, Präprozessoren oder Präprozessor-Softwaremodulen **212** zum Parsen von Strömen der Akustikdaten der Sprache in parametrische Repräsentationen wie beispielswei-

se Akustikmerkmale; einem oder mehreren Decodierern oder Decodierersoftwaremodulen **214** zum Decodieren der Akustikmerkmale, um daraus digitale Subwort- oder Wort-Ausgabedaten zu liefern, die den eingegebenen Sprachäußerungen entsprechen; und einen oder mehrere Backend-Prozessoren, Postprozessoren oder Postprozessor-Softwaremodule **216** zum Verwenden der von dem Decodierermodul/den Decodierermodulen **214** (nachfolgend numerusneutral 'Decodierermodul' genannt) gelieferten Ausgabedaten für einen beliebigen geeigneten Zweck.

[0033] Das System **210** kann außerdem Sprachmaterial von (einer) beliebigen geeigneten Audioquelle(n) **31** empfangen, die direkt mit dem Präprozessor-Softwaremodul/den Präprozessor-Softwaremodulen **212** (nachfolgend numerusneutral 'Präprozessor-Softwaremodul' genannt) in Verbindung gesetzt werden kann/können, wie in durchgehender Linie gezeigt, oder mit diesem/diesen indirekt über die Akustikschnittstelle **33** in Verbindung gesetzt werden kann/können. Bei der/den Audioquelle(n) **31** kann es sich zum Beispiel um eine Telefon-Audioquelle, wie etwa ein VoiceMail-System, oder um andere Telefondienste von beliebiger Art handeln.

[0034] Es können ein oder mehrere Module oder Modelle als Eingang in das Decodierermodul **214** verwendet werden. Erstens kann ein Grammatik- und/oder Lexikonmodell bzw. können Grammatik- und/oder Lexikonmodelle **218** Regeln bereitstellen, die festlegen, welche Wörter logikgemäß auf andere Wörter folgen können, um gültige Sätze zu bilden. Im weitgefassten Sinn kann ein Lexikon oder eine Grammatik eine Gesamtheit an Vokabular definieren, das zu einem beliebigen gegebenen Zeitpunkt in einem beliebigen gegebenen ASE-Modus von dem System **210** erwartet wird. Wenn sich das System **210** beispielsweise in einem Trainingsmodus zum Trainieren von Befehlen befindet, kann das Lexikon- oder Grammatikmodell bzw. können die Lexikon- oder Grammatikmodelle **218** alle Befehle umfassen, die dem System **210** bekannt sind und von diesem verwendet werden. Bei einem anderen Beispiel kann das aktive Lexikon- oder Grammatikmodell bzw. können die aktiven Lexikon- oder Grammatikmodelle **218**, wenn sich das System **210** in einem Hauptmenümodus befindet, alle Hauptmenübefehle umfassen, die von dem System **210** erwartet werden, wie beispielsweise Anrufen, Wählen, Verlassen, Löschen, Verzeichnis oder dergleichen. Zweitens wirkt ein Akustikmodell bzw. wirken Akustikmodelle **220** unterstützend bei der Auswahl der wahrscheinlichsten Subwörter oder Wörter, die einer von dem Präprozessormodul **212** kommenden Eingabe entsprechen. Drittens werden durch ein Wortmodell bzw. durch Wortmodelle **222** und durch ein Satz-/Sprachmodell bzw. durch Satz-/Sprachmodelle **224** Regeln, eine Syntax und/oder eine Semantik für das angemessene Platzieren der ausgewählten Subwörter oder Wörter in dem jeweils entsprechenden Wort- oder Satzkontext bereitgestellt. Außerdem kann durch das Satz-/Sprachmodell bzw. durch die Satz-/Sprachmodelle **224** eine Gesamtheit von Sätzen definiert werden, die das System **210** zu einem beliebigen gegebenen Zeitpunkt in einem beliebigen gegebenen ASE-Modus erwartet, und/oder können durch diese(s) Regeln etc. bereitgestellt werden, die bestimmen, welche Sätze logikgemäß auf andere Sätze folgen können, um einen gültigen erweiterten Sprachfluss zu bilden.

[0035] Gemäß einer alternativen beispielhaften Ausführungsform kann ein Teil oder die Gesamtheit des ASE-Systems **210** auf einer Rechenausrüstung installiert sein und mittels dieser verarbeitet werden, die sich an einem von dem Fahrzeug **12** entfernten Ort, wie beispielsweise dem Callcenter **20**, befindet. Beispielsweise können Grammatikmodelle, Akustikmodelle und dergleichen in dem Speicher eines der Server **82** und/oder einer der Datenbanken **84** in dem Callcenter **20** abgespeichert sein und zur fahrzeuginternen Sprachverarbeitung an die Fahrzeugtelematikeinheit **30** übermittelt werden. In ähnlicher Weise kann die Spracherkennungssoftware unter Verwendung von Prozessoren von einem der Server **82** in dem Callcenter **20** verarbeitet werden. Mit anderen Worten kann das ASE-System **210** in der Telematikeinheit **30** fest abgespeichert sein oder in beliebiger gewünschter Weise über das Callcenter **20** und das Fahrzeug **12** hinweg verteilt sein.

[0036] Als erstes werden aus menschlicher Sprache Akustikdaten extrahiert, wobei ein Fahrzeuginsasse in das Mikrophon **32** spricht, welches die Äußerungen in elektrische Signale umwandelt und solche Signale an die Akustikschnittstelle **33** übermittelt. Ein lautempfindliches Element in dem Mikrophon **32** erfasst die Sprachäußerungen des Insassen als Schwankungen des Luftdrucks und wandelt die Äußerungen in entsprechende Schwankungen analoger elektrischer Signale, wie beispielsweise von Gleichstrom oder Gleichspannung, um. Die Akustikschnittstelle **33** empfängt die analogen elektrischen Signale, die als Erstes derart abgetastet werden, dass Werte des Analogsignals zu diskreten Zeitpunkten erfasst werden, und anschließend derart quantisiert werden, dass die Amplituden der Analogsignale zu jedem Abtastzeitpunkt in einen kontinuierlichen Strom von digitalen Sprechdaten umgewandelt werden. Mit anderen Worten, die Akustikschnittstelle **33** wandelt die analogen elektrischen Signale in digitale elektronische Signale um. Bei den digitalen Daten handelt es sich um Binärbits, die in dem Telematikspeicher **54** zwischengespeichert und dann durch den Telematikprozessor **52** verarbeitet werden, oder die gleich zu Beginn, wenn sie empfangen werden, durch den Prozessor **52** in Echtzeit verarbeitet werden können.

[0037] Als Zweites wird durch das Präprozessormodul **212** der kontinuierliche Strom von digitalen

Sprechdaten in diskrete Abfolgen von Akustikparametern transformiert. Im Spezielleren führt der Prozessor **52** das Präprozessormodul **212** aus, um die digitalen Sprechdaten in einander überschneidende phonetische oder akustische Frames mit einer Dauer von beispielsweise 10–30 ms zu segmentieren. Die Frames entsprechen akustischen Subwörtern, wie beispielsweise Silben, Halbsilben, Phonen, Diphonen, Phonemen oder dergleichen. Durch das Präprozessormodul **212** erfolgt auch eine phonetische Analyse, um innerhalb jedes Frames Akustikparameter, wie beispielsweise zeitvariable Merkmalsvektoren, aus dem Sprachfluss des Insassen zu extrahieren. Äußerungen innerhalb des Sprachflusses des Insassen lassen sich als Abfolgen dieser Merkmalsvektoren darstellen. Es können beispielsweise Merkmalsvektoren in einer dem Fachmann bekannten Art und Weise extrahiert werden und diese können beispielsweise Tonhöhe, Energieprofile, Spektrum-Attribute und/oder Cepstrum-Koeffizienten umfassen, die gewonnen werden können, indem die Frames Fourier-Transformationen unterzogen werden und Akustikspektren unter Verwendung von Kosinustransformationen dekorreliert werden. Akustische Frames und entsprechende Parameter, die eine bestimmte Sprechdauer abdecken, werden zu unbekannten Testmustern von zu decodierendem Sprachfluss verknüpft.

[0038] Als Drittes führt der Prozessor das Decodierermodul **214** aus, um die eingehenden Merkmalsvektoren eines jeden Testmusters zu verarbeiten. Das Decodierermodul **214** wird auch Erkennungsmaschine oder Klassifizierer genannt und verwendet gespeicherte, bekannte Sprachfluss-Referenzmuster. Gleich den Testmustern, sind auch die Referenzmuster jeweils als eine Verknüpfung von miteinander in Beziehung stehenden akustischen Frames und entsprechenden Parametern definiert. Das Decodierermodul **214** vergleicht die Akustikmerkmalsvektoren eines zu erkennenden Subworttestmusters mit gespeicherten Subwortreferenzmustern und stellt diese gegenüber, bewertet den Umfang der Unterschiede oder Ähnlichkeiten zwischen ihnen und verwendet schließlich eine Entscheidungslogik, um ein am besten übereinstimmendes Subwort als das erkannte Subwort auszuwählen. Im Allgemeinen ist das am besten übereinstimmende Subwort jenes, welches dem gespeicherten, bekannten Referenzmuster entspricht, das eine minimale Unähnlichkeit zu dem Testmuster aufweist oder dieses mit der größten Wahrscheinlichkeit darstellt, wie es durch ein beliebiges von verschiedenen dem Fachmann bekannten Verfahren zum Analysieren und Erkennen von Subwörtern, bestimmt wird. Solche Techniken können 'Dynamit Time Warping'-Klassifizierer, Techniken aus der Künstlichen Intelligenz, neuronale Netze, Einrichtungen zur Erkennung freier Phoneme und/oder probabilistische Einrichtungen zur Inübereinstimmungbringung von Mustern, wie beispielsweise 'Hidden Markov Model'-Maschinen (HMM-Maschinen) umfassen.

[0039] HMM-Maschinen erzeugen, wie dem Fachmann bekannt, mehrere Hypothesen für Modelle zur Spracherkennung aus akustischen Eingabedaten. Die Hypothesen werden über eine Merkmalsanalyse des Sprachflusses beim abschließenden Identifizieren und Auswählen jener Erkennungsausgabedaten berücksichtigt, welche die am wahrscheinlichsten korrekte Decodierung der Akustikeingabedaten darstellen. Im Spezielleren erzeugt eine HMM-Maschine statistische Modelle in Form einer Liste der auf HMM-berechneten Konfidenzwerten oder Wahrscheinlichkeiten beruhenden ”N besten” Hypothesen für Subwortmodelle zu einer beobachteten Abfolge von Akustikdaten, der beispielsweise durch die Anwendung des Bayes'schen Theorems das eine oder andere Subwort zugewiesen wird.

[0040] Ein Bayes'scher HMM-Prozess identifiziert eine beste Hypothese, die der wahrscheinlichsten Äußerung oder Subwortabfolge für eine gegebene, beobachtete Abfolge von Akustikmerkmalsvektoren entspricht, und ihre Konfidenzwerte können von einer Vielzahl von Faktoren, einschließlich der mit den eingehenden Akustikdaten in Verbindung stehenden, akustischen Signal/Rauschverhältnisse, abhängen. Das HMM kann auch eine statistische Verteilung umfassen, welche eine Mischung von diagonalen Gaußverteilungen genannt wird, die einen Wahrscheinlichkeits-Punktewert für jeden beobachteten Merkmalsvektor jedes Subworts liefert, wobei die Punktewerte dazu verwendet werden können, die Liste der N besten Hypothesen neu zu ordnen. Die HMM-Maschine kann auch ein Subwort identifizieren und auswählen, dessen Wahrscheinlichkeitspunktewert für die Modellentsprechung am höchsten ist.

[0041] In ähnlicher Weise können einzelne HMMs für eine Abfolge von Subwörtern miteinander verknüpft werden, um Einwort- oder Mehrwort-HMMs zu erstellen. Daraufhin kann eine Liste der N besten Einwort- oder Mehrwort-Referenzmuster samt zugeordneten Parameterwerten generiert und weiter ausgewertet werden.

[0042] Bei einem Beispiel verarbeitet der Spracherkennungsdecodierer **214** die Merkmalsvektoren unter Verwendung der geeigneten Akustikmodelle, Grammatiken und Algorithmen, um eine Liste der N besten Referenzmuster zu erzeugen. Der hier verwendete Begriff Referenzmuster ist durch Modelle, Wellenformen, Vorlagen, Rich-Signal-Modelle, Musterbeispiele, Hypothesen oder andere Referenztypen austauschbar. Ein Referenzmuster kann eine Reihe von Merkmalsvektoren umfassen, die ein oder mehrere Wörter oder Subwörter repräsentieren, und kann auf individuellen Sprechern, Sprecharten und Hörumfeldbedingungen basieren. Der Fachmann wird er-

kennen, dass Referenzmuster durch ein geeignetes Referenzmustertraining des ASE-Systems erzeugt und abgespeichert werden können. Der Fachmann wird außerdem erkennen, dass abgespeicherte Referenzmuster bearbeitet werden können, wobei Parameterwerte der Referenzmuster auf der Grundlage von Unterschieden bei dem Spracheingangssignal zwischen dem Referenzmustertraining und der tatsächlichen Verwendung des ASE-Systems angepasst werden. Beispielsweise kann ein Satz von Referenzmustern, die für einen bestimmten Fahrzeuginsassen oder für bestimmte Akustikbedingungen trainiert worden sind, angepasst werden und als ein anderer Satz von Referenzmustern für einen anderen Fahrzeuginsassen oder für andere Akustikbedingungen abgespeichert werden, und zwar auf der Grundlage einer begrenzten Menge von Trainingsdaten, die jeweils von dem anderen Fahrzeuginsassen oder den anderen Akustikbedingungen stammen. Mit anderen Worten, die Referenzmuster sind nicht notwendigerweise feststehend und können während der Spracherkennung nachjustiert werden.

[0043] Unter Verwendung der vokabularinternen Grammatik und eines beliebigen geeigneten Decodiereralgorithmus und Akustikmodells/beliebiger geeigneter Decodiereralgorithmen und Akustikmodelle (nachfolgend numerusneutral 'Decodiereralgorithmus' bzw. 'Akustikmodell' genannt) ruft der Prozessor von dem Speicher verschiedene Referenzmuster ab, durch welche das Testmuster interpretiert wird. Beispielsweise kann der Prozessor eine Liste der N besten Vokabularergebnisse oder Referenzmuster erzeugen und zusammen mit entsprechenden Parameterwerten abspeichern. Beispielhafte Parameterwerte können Konfidenzpunktewerte eines jeden Referenzmusters in der Liste der N besten Vokabularergebnisse und zugehörige Segmentdauern, Wahrscheinlichkeits-Punktewerte, Signal/Rauschverhältniswerte (SNR-Werte) und/oder dergleichen umfassen. Die Liste der N besten Vokabularergebnisse kann nach der Parameterwertgröße absteigend angeordnet werden. Beispielsweise ist das Referenzmuster aus den Vokabularergebnissen, welches den höchsten Konfidenzpunktewert aufweist, das erste der N besten Referenzmuster und so weiter. Sobald eine Kette von erkannten Subwörtern erstellt ist, können diese dazu verwendet werden, unter Zuhilfenahme von Eingangsdaten aus den Wortmodellen **222** Wörter zu konstruieren und unter Zuhilfenahme von Eingangsdaten aus den Sprachmodellen **224** Sätze zu konstruieren.

[0044] Schließlich empfängt das Postprozessor-Softwaremodul/empfangen die Postprozessor-Softwaremodule **216** (nachfolgend numerusneutral 'Postprozessor-Softwaremodul' genannt) die Ausgangsdaten von dem Decodierermodul **214** im Hinblick auf einen beliebigen geeigneten Zweck. Bei einem Beispiel kann das Postprozessor-Softwaremodul **216** eines der Referenzmuster aus der Liste der N besten Einwort- oder Mehrwort-Referenzmuster als erkannten Sprachfluss identifizieren oder auswählen. Bei einem anderen Beispiel kann das Postprozessormodul **216** dazu verwendet werden, Akustikdaten in Text oder Ziffern zur Verwendung mit anderen Aspekten des ASE-Systems oder mit anderen Fahrzeugsystemen umzuwandeln. Bei einem weiteren Beispiel kann das Postprozessormodul **216** dazu verwendet werden, dem Decodierer **214** oder dem Präprozessor **212** eine Trainingsrückmeldung zu liefern. Im Spezielleren kann der Postprozessor **216** dazu verwendet werden, Akustikmodelle für das Decodierermodul **214** zu trainieren, oder Abstimmungsparameter für das Präprozessormodul **212** zu trainieren.

Verfahren

[0045] In Fig. 3 und Fig. 4, die nun einer genaueren Betrachtung unterzogen werden, sind spracherkennungsbezogene Verfahren **300**, **400** gezeigt, die unter Verwendung einer geeigneten Programmierung des ASE-Systems **210** aus Fig. 2 innerhalb der Betriebsumgebung der Fahrzeugtelematikeinheit **30** sowie unter Verwendung geeigneter Hardware und unter geeigneter Programmierung der anderen in Fig. 1 gezeigten Komponenten realisiert werden können. Eine solche Programmierung und Verwendung der weiter oben beschriebenen Hardware wird für den Fachmann auf der Grundlage der obigen Systembeschreibung und der Erörterung des weiter unten in Verbindung mit den verbleibenden Figuren beschriebenen Verfahrens ersichtlich. Für den Fachmann ist außerdem erkenntlich, dass die Verfahren auch unter Verwendung anderer ASE-Systeme innerhalb anderer Betriebsumgebungen realisiert werden können.

[0046] Im Allgemeinen wird durch ein Verfahren **300** zur Erzeugung von Akustikmodellen die automatische Spracherkennung verbessert, indem Akustikmodelle, die ausgehend von Sprachmaterial trainiert worden sind, das von männlichen Sprechern einer ersten Sprache gewonnen worden ist, in Reaktion auf sprachenunabhängige Sprechdaten von weiblichen Sprechern einer zweiten Sprache angepasst werden. Demgemäß werden Proxy-Akustikmodelle zur Verwendung während der Laufzeit der Spracherkennung einer Äußerung eines weiblichen Sprechers der ersten Sprache erzeugt. Die Erkennung des Sprachflusses eines weiblichen Sprechers wird verbessert, weil ungeachtet der Tatsache, dass für bestimmte weibliche Sprecher keine Trainingsdaten für Akustikmodelle verfügbar sind, gute weibliche Proxy-Akustikmodelle einer gegebenen Sprache verwendet werden können.

[0047] In Fig. 3, auf welche nun Bezug genommen wird, beginnt das Verfahren **300** in einer beliebigen, geeigneten Weise bei Schritt **305**.

[0048] In Schritt **310** können Akustikmodelle unter Heranziehung von männlichen Sprechern einer ersten Sprache trainiert werden. Das Training von Akustikmodellen kann Schätzverfahren der maximalen Wahrscheinlichkeit (Maximum-Likelihood-Methode), Verfahren der minimalen Klassifizierungsfehler, konditionale Maximum-Likelihood-Verfahren, welche alle dem Fachmann allgemein bekannt sind, umfassen und es können auch beliebige andere geeignete Verfahren verwendet werden. Bei den Akustikmodellen kann es sich um Wortmodelle, Phonemmodelle oder Subwortmodelle handeln. Jede geeignete Anzahl von männlichen Sprechern kann dafür herangezogen werden. Bei der ersten Sprache kann es sich um Arabisch, Chinesisch oder jede beliebige andere Sprache handeln.

[0049] In Schritt **315** werden die in Schritt **310** trainierten Akustikmodelle in Reaktion auf sprachenunabhängige Sprechdaten von weiblichen Sprechern einer zweiten Sprache angepasst, um Proxy-Akustikmodelle zur Verwendung während der Laufzeit der Spracherkennung einer Äußerung eines weiblichen Sprechers der ersten Sprache zu erzeugen. Beispielsweise können Maximum-Likelihood-Linear-Regression-Verfahren (MLLR-Verfahren) Sprechdaten von weiblichen Sprechern des Englischen dazu verwenden, die Mittelfrequenzen von Akustikmodellen von männlichen Sprechern des Arabischen nachzujustieren. Beliebige geeignete MLLR-Verfahren können dazu verwendet werden und sind dem Fachmann auch allgemein bekannt, wie aus dem Fachartikel 'Variance compensation within the MLLR framework for robust speech recognition and speaker adaptation (Varianzkompensation innerhalb der MLLR-Umgebung im Hinblick auf eine robuste Spracherkennung und Sprecheranpassung), Gales, M., D. Pye, und P. Woodland, In Proc. ICSLP, pp. 1832–1835, (1996), hervorgeht.

[0050] In Schritt **320** können die in Schritt **315** erzeugten Proxy-Akustikmodelle während der Laufzeit der automatischen Spracherkennung zum Erkennen der Sprache eines weiblichen Sprechers der ersten Sprache verwendet werden.

[0051] In **Fig.** 325 kann das Verfahren in einer beliebigen geeigneten Weise enden.

[0052] In einer Ausführungsform des Schrittes **315** werden die Akustikmodelle vor der Laufzeit der Spracherkennung angepasst. Es können beispielsweise Maximum-Likelihood-Linear-Regression-Verfahren oder beliebige andere geeignete Anpassungsverfahren verwendet werden, um im Rahmen der Anwendung einer Vorlage für Formantenfrequenzen und Frequenzbänder sprachenunabhängiger Sprechdaten die Mittelfrequenzen der Akustikmodelle derart anzupassen, dass diese die weiblichen Proxy-Akustikmodelle zum Ergebnis haben.

[0053] In einer anderen Ausführungsform des Schrittes **315** werden die Akustikmodelle während der Laufzeit des Sprachflusses angepasst. Diese Ausführungsform ist weiter unten unter Bezugnahme auf Fig. 4 und Verfahren **400** im Einzelnen beschrieben.

[0054] Im Allgemeinen wird durch das Spracherkennungsverfahren **400** die Spracherkennung verbessert, indem Akustikmodelle, die ausgehend von Sprachmaterial von männlichen Sprechern einer ersten Sprache trainiert worden sind, in Reaktion auf das Identifizieren von zumindest einem aus einer Mehrzahl von Formantenfrequenzbändern in Sprechdaten von weiblichen Sprechern einer zweiten Sprache, das zumindest einer Formantenfrequenz in dem Sprachfluss entspricht, der von einem weiblichen Sprecher der ersten Sprache gewonnen wird, angepasst werden. Die Erkennung des Sprachflusses eines weiblichen Sprechers wird verbessert, weil ungeachtet der Tatsache, dass für bestimmte weibliche Sprecher keine Trainingsdaten für Akustikmodelle verfügbar sind, gute weibliche Proxy-Akustikmodelle einer gegebenen Sprache verwendet werden können.

[0055] In Fig. 4, auf welche nun Bezug genommen wird, beginnt das Verfahren **400** in einer beliebigen geeigneten Weise bei Schritt **405**. Ein Fahrzeugbenutzer startet beispielsweise die Interaktion mit der Benutzerschnittstelle der Telematikeinheit **30** vorzugsweise durch Drücken der Drucktaste **34** der Benutzerschnittstelle, um eine Sitzung zu beginnen, bei welcher der Benutzer Sprachbefehle eingibt, die von der Telematikeinheit **30** interpretiert werden, während diese im Spracherkennungsmodus betrieben wird.

[0056] Unter Verwendung des Audiosystems **36** kann die Telematikeinheit **30** die Betätigung der Drucktaste durch das Abspielen eines Tons oder durch die verbale Aufforderung zu einer Befehlseingabe seitens des Benutzers oder Insassen bestätigen. Das Verfahren **400** wird während der Laufzeit der Spracherkennung ausgeführt.

[0057] In Schritt **410** wird die Sprache in einer beliebigen geeigneten Weise empfangen. So kann etwa das Telematikmikrophon **32** Sprachäußerungen eines Benutzers empfangen und die Akustikschnittstelle **33** kann den Sprachfluss in Akustikdaten digitalisieren. In einer Ausführungsform handelt es sich bei der Sprachäußerung um einen Befehl, beispielsweise um einen in einem Systemmenü erwarteten Befehl. In einer spezielleren Ausführungsform handelt es sich bei dem Befehl um ein erstes Befehlswort in einem System-Hauptmenü nach Beginn des Verfahrens **300**. Der Sprachfluss kann eine oder mehrere Äußerungen umfassen.

[0058] In Schritt **415** wird der empfangene Sprachfluss vorverarbeitet, um Akustikmerkmalsvektoren zu erzeugen. Die von der Akustikschnittstelle **33** kommenden Akustikdaten können beispielsweise durch das Präprozessormodul **212** des ASE-Systems **210** wie weiter oben beschrieben vorverarbeitet werden.

[0059] In Schritt **420** werden sprachenunabhängige Sprechdaten des empfangenen Sprachflusses bestimmt. Beispielsweise kann die Tonhöhe und/oder eine oder mehrere Formantenfrequenzen des empfangenen Sprachflusses in beliebiger geeigneter Weise geschätzt, detektiert oder sonst wie bestimmt werden. In einem Beispiel stellt die Cepstrumanalyse eine gute Schätzung von Grundharmonischer, Tonhöhenfrequenz und dergleichen bereit, und liefert Informationen über glottale Pulse und Vokaltrakt-Frequenzgang. Als weitere Beispiele sind zu nennen die Analyse der Nulldurchgangsrate in der Zeitdomäne des Sprachsignals, Linearvorhersage-Codierungsalgorithmen und/oder Autokorrelationsanalyse des Sprachflusses in der Zeitdomäne. Solche Verfahren sind dem Fachmann allgemein bekannt, wie aus verschiedenen Veröffentlichungen hervorgeht, wobei hier zu nennen sind: Oppenheim, A. V., Johnson, D. H. (1972) 'Discrete representation of signals' (Diskrete Repräsentation von Signalen), Proceedings of the IEEE, Bd. 60, Nr. 6, pp. 681–691, June 1972; 'Speech and Audio Signal Processing, Processing & Perception of Speech and Music' (Sprach- und Audiosignalverarbeitung, Verarbeitung und Perzeption von Sprache und Musik), Ben Gold und Nelson Morgan, John Wiley & Sons, 2001; und 'Fundamentals of Speech Recognition' (Grundlagen der Spracherkennung), Lawrence Rabiner & Biing-Hwang Juang, Prentice Hall, 1993.

[0060] In Schritt **425** kann ein Spracherkennungs-Präprozessor sprachenunabhängige Sprechdaten von weiblichen Sprechern einer zweiten Sprache identifizieren, die den in Schritt **420** bestimmten sprachenunabhängigen Sprechdaten entsprechen. Die sprachenunabhängigen Sprechdaten können beispielsweise die Tonhöhe und/oder zumindest eines aus einer Mehrzahl von Formantenfrequenzbändern oder Bins aus dem Sprachfluss von weiblichen Sprechern der zweiten Sprache umfassen. Die Sprechdaten von weiblichen Sprechern der zweiten Sprache können Datenbanken mit voraufgezeichneten Sprachäußerungen, statistischen Modellen, einer oder mehreren Formantenfrequenzvorlagen der Sprachäußerung und/oder dergleichen enthalten. In einer spezifischen Veranschaulichung können die Sprechdaten eine Formantenfrequenzvorlagenverteilung mit 20 Bändern oder Bins von einander überlappenden Frequenzbereichen und beabstandeten Mittelfrequenzen umfassen. Erste Formantenfrequenzen in der Vorlage können sich von 600 Hz bis 1200 Hz über 12 der 20 Bins mit einem Abstand von 50 Hz erstrecken. In diesem Fall kann eine Äußerung des weiblichen Sprechers der ersten Sprache eine Formantenfrequenz von 900 Hz umfassen, welche dem achten Band oder Bin der Vorlage entspricht.

[0061] In Schritt **430** werden Akustikmodelle, die aus dem Sprachfluss von männlichen Sprechern der ersten Sprache gewonnen worden sind, in Reaktion auf die Identifizierung von sprachenunabhängigen Sprechdaten aus Schritt **425** angepasst. Die Anpassung von Akustikmodellen ist dem Fachmann allgemein bekannt und es können dazu beliebige geeignete Verfahren verwendet werden. Die Akustikmodelle können beispielsweise durch Frequenzverzerrung der Akustikmodelle mit Formantenfrequenzdaten von weiblichen Sprechern der ersten Sprache angepasst werden, so dass die Akustikmodelle im Hinblick auf eine bessere Übereinstimmung mit dem empfangenen Sprachfluss angepasst sind. Ein besonderes Verfahren, das hierzu verwendet werden kann, umfasst die sogenannte Vokaltraktlängennormalisierung (Vocal Tract Length Normalization, VTLN), wobei die Tonhöhe und/oder die Formantenfrequenzen für den weiblichen Sprecher der ersten Sprache zur Anpassung der männlichen Akustikmodelle verwendet werden. Die VTLN ist dem Fachmann allgemein bekannt, wie dies aus dem Fachartikel 'Vocal Tract Length Normalization in Speech Recognition-Compensation for Systematic Speaker Variability' (Vokaltraktlängennormalisierung bei der Spracherkennungs-Kompensation für systematische Sprechervariabilität), Kamm. T., Andreou A, Cohen J., Proceedings of 15th Annual Speech Research Symposium, Baltimore, MY, Juni 1995, hervorgeht.

[0062] In einer Ausführungsform kann die Formantenfrequenz des Bestimmungsschritts **420** einen Mittelwert aus einer Mehrzahl von Formantenfrequenzen der empfangenen Äußerung umfassen. Im Spezielleren kann die Mehrzahl von Formantenfrequenzen der empfangenen Äußerung einen ersten Formanten und/oder einen zweiten Formanten und/oder einen dritten Formanten umfassen. In einem Beispiel kann die erste Formantenfrequenz wie folgt charakterisiert sein: Mittlere F1 = F1(Konsonant1) + F1(Konsonant2) + F1(Vokal1) + F1(Konsonant3) ... F1(Nte Silbe)/N. Ähnliche Charakterisierungen können für die Formanten F2 und/oder F3 vorgenommen werden.

[0063] In einer anderen Ausführungsform umfasst die Formantenfrequenz des Bestimmungsschritts **420** einen ersten Formanten, einen zweiten Formanten und einen dritten Formanten. In diesem Beispiel werden die drei Formantenfrequenzen, nämlich F1, F2 und F3 dazu verwendet, die Mittelfrequenzen von Filterbändern in den männlichen Akustikmodellen nachzujustieren. Wenn F_{Mittel} = Funktion (F_1, F_2, F_3), dann ist $\lambda_{Transformiert}$ = Funktion ($\lambda_{Männlich}$, F_{Mittel}).

[0064] Unter Weiterführung des Beispiels aus Schritt **425** können beliebige geeignete Frequenzverzerrungsverfahren an den männlichen Akustikmodellen durchgeführt werden, die darauf basieren, dass die identifizierte Frequenz in einem speziellen Band oder Bin, z. B. in dem achten Band oder Bin, gelegen ist. Demzufolge können die Akustikmodelle bandspezifisch einer Frequenzverzerrung unterzogen werden.

[0065] In Schritt **435** werden die erzeugten Akustikmerkmalsvektoren unter Verwendung der angepassten Akustikmodelle aus Schritt **425** decodiert, um eine Mehrzahl von Hypothesen für den empfangenen Sprachfluss zu erstellen. Bei der Mehrzahl von Hypothesen kann es sich beispielsweise um eine Liste der N besten Hypothesen handeln und das Decodierermodul **214** des ASE-Systems **210** kann dazu verwendet werden, die Akustikmerkmalsvektoren zu decodieren.

[0066] In Schritt **440** wird die Mehrzahl von Hypothesen nachverarbeitet, um eine aus der Mehrzahl von Hypothesen als die empfangene Sprachäußerung zu identifizieren. Der Postprozessor **216** des ASE-Systems **210** kann beispielsweise die Hypothesen nachverarbeiten, um die erste der besten Hypothesen als die empfangene Sprachäußerung zu identifizieren. In einem anderen Beispiel kann der Postprozessor **216** die Liste der N besten Hypothesen in beliebiger geeigneter Weise neu anordnen und nach der Neuanordnung die erste der besten Hypothesen identifizieren.

[0067] In Schritt **445** kann das Verfahren **400** in einer beliebigen geeigneten Weise enden.

[0068] Die Verfahren oder Teile davon können in einem Computerprogrammprodukt implementiert sein, das Befehle umfasst, die auf einem computerlesbaren Trägermedium zur Verwendung durch einen oder mehrere Prozessoren von einem oder mehreren Computer vorhanden sind, um einen oder mehrere der Verfahrensschritte zu implementieren. Das Computerprogrammprodukt kann umfassen: ein oder mehrere Softwareprogramme, die aus Programmbefehlen in Quellcode, Objektcode, ausführbarem Code oder in anderen Formaten bestehen; ein oder mehrere Firmwareprogramme; oder Dateien in Hardwarebeschreibungssprache (HDL-Dateien); sowie beliebige sonstige programmbezogene Daten. Die Daten können Datenstrukturen, Nachschlagetabellen oder Daten in einem beliebigen sonstigen, geeigneten Format umfassen. Die Programmbefehle können Programmmodule, Routinen, Programme, Objekte, Komponenten und/oder dergleichen umfassen. Das Computerprogramm kann auf einem Computer oder auf mehreren miteinander kommunizierenden Computer ausgeführt werden.

[0069] Das Programm kann/Die Programme können auf einem computerlesbaren Medium vorhanden sein, welches eine oder mehrere Speichereinrichtungen, Fertigungsartikel oder dergleichen umfassen kann. Als beispielhafte computerlesbare Medien sind zu nennen: Computer-Systemspeicher, z. B. RAM (Direktzugriffsspeicher), ROM (Festwertspeicher); Halbleiterspeicher, z. B. EPROM (lösch- und programmierbarer ROM), EEPROM (elektrisch löschbarer und programmierbarer ROM), Flash-Speicher; magnetische oder optische Speicherplatten oder Bänder; und/oder dergleichen. Das computerlesbare Medium kann auch Verbindungen zwischen Computern umfassen, beispielsweise wenn Daten über ein Netzwerk oder eine andere Kommunikationsverbindung (entweder drahtgebunden, drahtlos oder kombiniert) übertragen oder bereitgestellt werden. Jegliche beliebige(n) Kombination(en) aus den oben beschriebenen Beispielen ist/sind ebenfalls im Umfang des computerlesbaren Mediums enthalten. Es versteht sich daher, dass das Verfahren zumindest teilweise durch alle beliebigen Elektronikartikel und/oder -einrichtungen ausgeführt werden kann, die in der Lage sind, Befehle auszuführen, welche einem oder mehreren Schritten des offenbarten Verfahrens entsprechen.

[0070] Es versteht sich, dass es sich bei dem weiter oben Erwähnten um eine Beschreibung einer oder mehrerer bevorzugter, beispielhafter Ausführungsformen der Erfindung handelt. Die Erfindung ist nicht auf die hier offenbarte(n) spezielle(n) Ausführungsform(en) beschränkt, sondern ist vielmehr ausschließlich durch die nachfolgenden Ansprüche definiert.

[0071] Darüber hinaus beziehen sich die in der obigen Beschreibung getroffenen Aussagen auf besondere Ausführungsformen und sind nicht als Einschränkungen des Umfangs der Erfindung oder der Definition von Begriffen, die in den Ansprüchen verwendet werden, auszulegen, es sei denn, ein Begriff oder eine Phrase ist weiter oben ausdrücklich definiert. Verschiedene andere Ausführungsformen und verschiedene Abänderungen und Modifikationen der offenbarten Ausführungsform(en) sind für den Fachmann ersichtlich. Die Erfindung kann beispielsweise auf andere Gebiete der Sprachsignalverarbeitung, wie etwa auf die mobile Telekommunikation, auf Internettelefonie-Anwendungen (VoIP-Anwendungen) und dergleichen angewendet werden. Alle solchen anderen Ausführungsformen, Veränderungen und Modifikationen sollen innerhalb des Umfangs der beigefügten Ansprüche gelegen sein.

[0072] Die in dieser Beschreibung und in den Ansprüchen verwendeten Begriffe ”beispielsweise”, ”zum Beispiel”, ”wie zum Beispiel” und ”wie”, sowie die Verben ”umfassend”, ”aufweisend”, ”einschließend” und deren andere Verbformen in Verbindung

mit einer Auflistung eines oder mehrerer Bauteile oder sonstiger Gegenstände, sind jeweils als unbeschränkt und offen aufzufassen, was bedeutet, dass die Auflistung nicht so zu betrachten ist, dass andere, zusätzliche Bauteile oder Gegenstände davon ausgeschlossen sind. Andere Begriffe sind in ihrer weitesten, vernünftigen Bedeutung aufzufassen, sofern sie nicht in einem Zusammenhang verwendet sind, der eine andere Auslegung erfordert.

Patentansprüche

1. Verfahren zum Erzeugen von Proxy-Akustikmodellen für die Verwendung bei der automatischen Spracherkennung, welches die Schritte umfasst, dass:
(a) ausgehend von Sprachmaterial, das über Mikrophon von männlichen Sprechern einer ersten Sprache gewonnen worden ist, Akustikmodelle trainiert werden, und
(b) die in Schritt (a) trainierten Akustikmodelle in Reaktion auf sprachenunabhängige Sprechdaten von weiblichen Sprechern einer zweiten Sprache angepasst werden, um Proxy-Akustikmodelle zur Verwendung während der Laufzeit der Spracherkennung einer Äußerung eines weiblichen Sprechers der ersten Sprache zu erzeugen.

2. Verfahren nach Anspruch 1, wobei der Anpassungsschritt (b) vor der Laufzeit der Spracherkennung durchgeführt wird.

3. Verfahren nach Anspruch 1, wobei der Anpassungsschritt (b) an der Äußerung des weiblichen Sprechers der ersten Sprache während der Laufzeit der Spracherkennung durchgeführt wird.

4. Verfahren nach Anspruch 3, wobei der Anpassungsschritt (b) in Reaktion auf eine Identifizierung von zumindest einem aus einer Mehrzahl von Formantenfrequenzbändern in den Sprechdaten der weiblichen Sprecher der zweiten Sprache erfolgt, das zumindest einer Formantenfrequenz entspricht, die in der Äußerung des weiblichen Sprechers der ersten Sprache bestimmt wird.

5. Verfahren nach Anspruch 4, wobei der Anpassungsschritt (b) durch Frequenzverzerrung der in Schritt (a) trainierten Akustikmodelle in Reaktion auf die Identifizierung des zumindest einen aus der Mehrzahl von Formantenfrequenzbändern in den Sprechdaten der weiblichen Sprecher der zweiten Sprache durchgeführt wird.

6. Verfahren nach Anspruch 4, wobei es sich bei der zumindest einen Formantenfrequenz, die in der Äußerung des weiblichen Sprechers der ersten Sprache bestimmt wird, um einen Mittelwert aus einer Mehrzahl von Formantenfrequenzen in der empfangenen Äußerung handelt.

7. Verfahren nach Anspruch 6, wobei die Mehrzahl von Formantenfrequenzen in der empfangenen Äußerung von einem ersten Formanten und/oder einem zweiten Formanten und/oder einem dritten Formanten stammt.

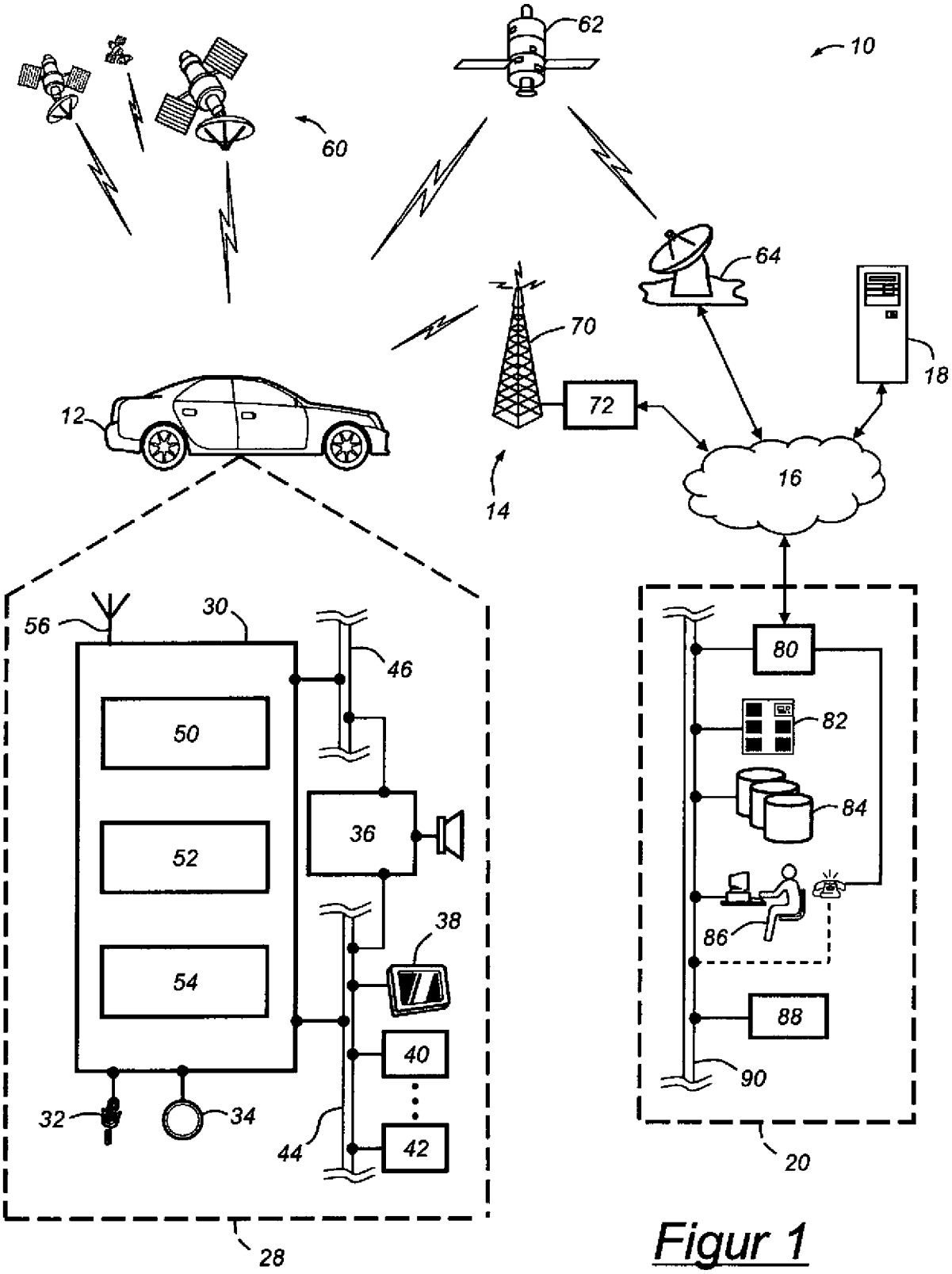
8. Verfahren nach Anspruch 6, wobei die zumindest eine Formantenfrequenz des Bestimmungsschritts (c) einen ersten Formanten, einen zweiten Formanten und einen dritten Formanten umfasst.

9. Verfahren zur automatischen Spracherkennung, welches die Schritte umfasst, dass:
(a) eine Äußerung über ein Mikrophon von einem weiblichen Sprecher einer ersten Sprache empfangen wird;
(b) die Äußerung mit einem Präprozessor zur automatischen Spracherkennung vorverarbeitet wird, um Akustikmerkmalsvektoren zu erzeugen;
(c) zumindest eine Formantenfrequenz der empfangenen Äußerung bestimmt wird;
(d) zumindest eines aus einer Mehrzahl von Formantenfrequenzbändern in Sprechdaten von weiblichen Sprechern einer zweiten Sprache identifiziert wird, das der zumindest einen in Schritt (c) bestimmten Formantenfrequenz entspricht; und
(e) Akustikmodelle, die ausgehend von Sprachmaterial von männlichen Sprechern der ersten Sprache trainiert worden sind, in Reaktion auf den Identifikationsschritt (d) angepasst werden, um als Ergebnis Proxy-Akustikmodelle für die weiblichen Sprecher der ersten Sprache zu liefern.

10. Verfahren nach Anspruch 9, welches weiterhin umfasst, dass als Schritt (f) die in Schritt (b) erzeugten Akustikmerkmalsvektoren unter Verwendung eines Prozessors und der in Schritt (e) angepassten Akustikmodelle decodiert werden, um eine Mehrzahl von Hypothesen für die empfangene Äußerung zu erstellen.

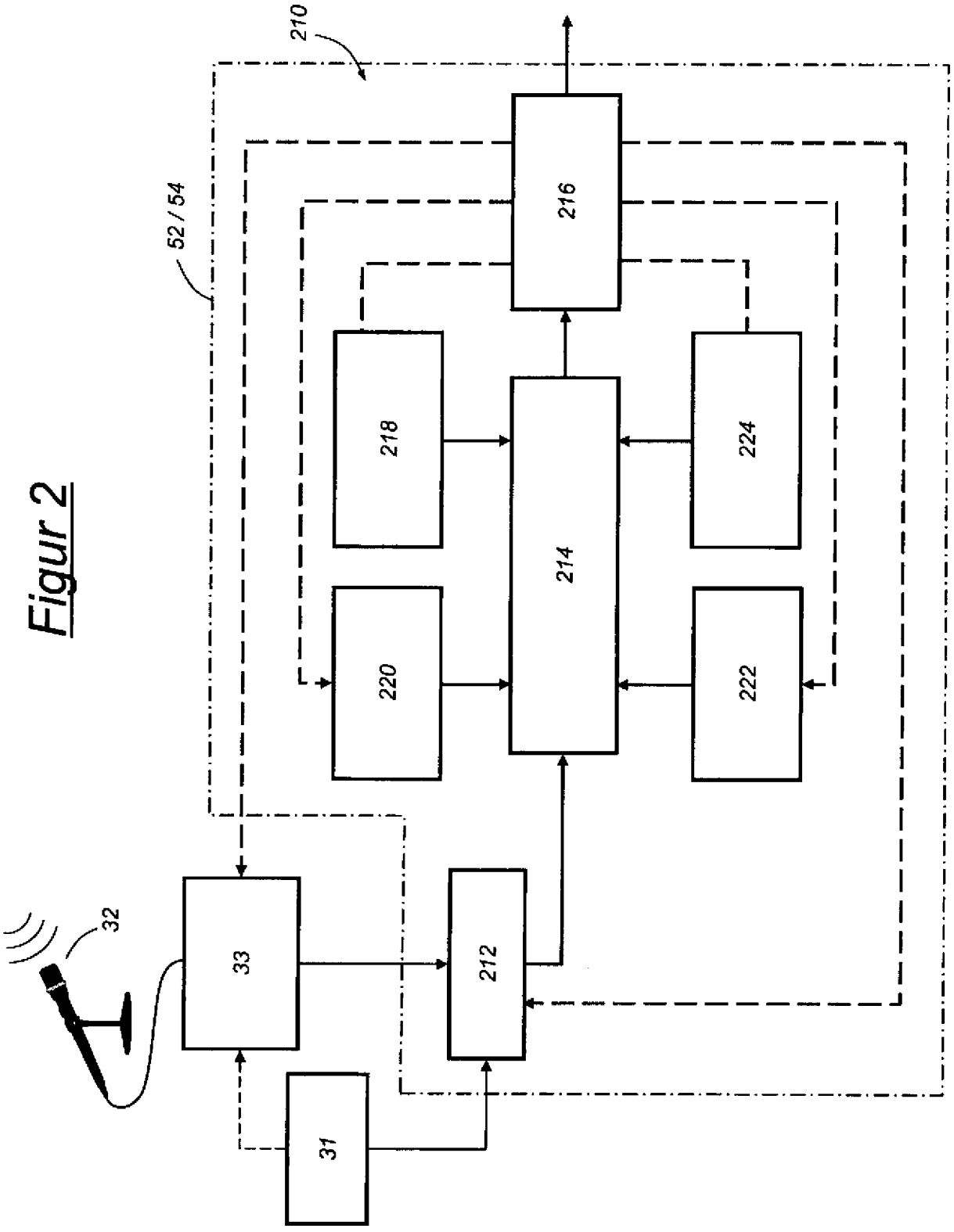
Es folgen 3 Blatt Zeichnungen

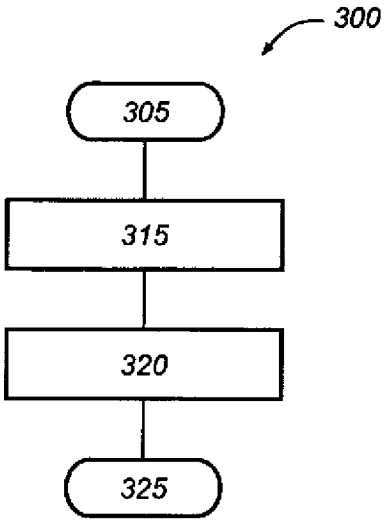
Anhängende Zeichnungen



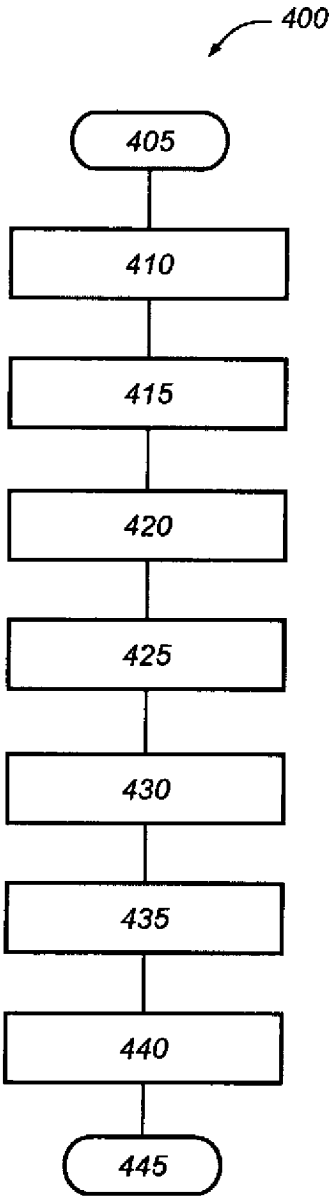
Figur 1

Figur 2





Figur 3



Figur 4