



(10) **DE 10 2014 203 540 A1** 2015.02.05

(12)

Offenlegungsschrift

(21) Aktenzeichen: **10 2014 203 540.6**

(22) Anmeldetag: **27.02.2014**

(43) Offenlegungstag: **05.02.2015**

(51) Int Cl.: **G10L 15/22 (2006.01)**

(30) Unionspriorität:

13/955,579 **31.07.2013** **US**

(71) Anmelder:

**GM GLOBAL TECHNOLOGY OPERATIONS, LLC,
Detroit, Mich., US**

(74) Vertreter:

**isarpatent Patentanwälte Behnisch, Barth,
Charles, Hassa, Peckmann & Partner mbB, 80801
München, DE**

(72) Erfinder:

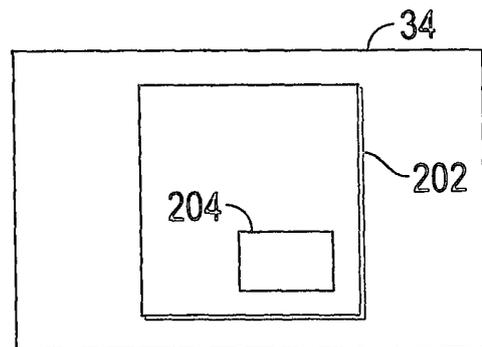
**Tzirkel-Hancock, Eli, Ra'Anana, IL; Sims, Robert
D., Milford, Mich., US; Tsimhoni, Omer, Ramat
Hasharon, IL**

Prüfungsantrag gemäß § 44 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

(54) Bezeichnung: **SYSTEME UND VERFAHREN FÜR DAS STEuern EINES DIALOG-KONTEXTES IN
SPRACHSYSTEMEN**

(57) Zusammenfassung: Es werden Verfahren und Systeme für das Steuern eines gesprochenen Dialoges innerhalb eines Sprachsystems bereitgestellt. Das Verfahren beinhaltet das Erstellen einer gesprochenen Dialog-Sitzung, welche einen ersten Dialog-Kontext besitzt, und das Empfangen eines Kontext-Triggers, welcher zu einer Aktion gehört, welche von einem Benutzer durchgeführt wird. In Antwort auf den Kontext-Trigger wechselt das System zu einem zweiten Dialog-Kontext. In Antwort auf einen Kontext-Vollendungszustand kehrt das System zu dem ersten Dialog-Kontext zurück.



Beschreibung**ZUSAMMENFASSUNG****TECHNISCHER BEREICH**

[0001] Der technische Bereich bezieht sich im Allgemeinen auf Sprachsysteme, und spezieller ausgedrückt bezieht er sich auf Verfahren und Systeme für das Steuern eines Dialog-Kontextes innerhalb eines Sprachsystems.

HINTERGRUND

[0002] Gesprochene Dialogsysteme oder "Sprachsysteme" in Fahrzeugen führen, neben anderen Dingen, Spracherkennung durch, basierend auf einer Sprache, welche durch Insassen des Fahrzeugs geäußert wird. Die Sprachäußerungen beinhalten typischerweise Befehle, welche mit einem oder mit mehreren Merkmalen des Fahrzeugs ebenso wie mit anderen Systemen, welche für das Fahrzeug zugänglich sind, kommunizieren oder diese steuern. Ein Sprachsystem erzeugt gesprochene Befehle in Antwort auf die Sprachäußerungen, und in einigen Fällen werden die gesprochenen Befehle in Antwort auf die Spracherkennung erzeugt, welche weitere Information benötigt, um die Spracherkennung durchzuführen.

[0003] In vielen Fällen kann der Benutzer wünschen, das gesprochene Dialogthema zu ändern, bevor die Sitzung vollendet ist. D. h., der Benutzer kann wünschen, den "Dialog-Kontext" während einer Sitzung zu ändern. Dies kann zum Beispiel auftreten, wenn: (1) der Benutzer weitere Information benötigt, um die Aufgabe zu vollenden, (2) der Benutzer eine Aufgabe nicht beenden kann, (3) der Benutzer seine Absicht oder ihre Absicht geändert hat, (4) das Sprachsystem einen falschen Pfad in dem gesprochenen Dialog eingenommen hat oder (5) der Benutzer unterbrochen wurde. In gegenwärtig bekannten Systemen führen derartige Szenarien häufig zu einem Dialog-Fehler und zur Frustration des Benutzers. Zum Beispiel kann der Benutzer die erste gesprochene Dialog-Sitzung aufgeben, eine neue gesprochene Dialog-Sitzung beginnen, um fehlende Information zu bestimmen, und dann noch eine andere gesprochene Dialog-Sitzung beginnen, um die ursprüngliche Aufgabe, welche für die erste Sitzung gedacht ist zu vollenden.

[0004] Entsprechend ist es wünschenswert, verbesserte Verfahren und Systeme für das Steuern eines Dialog-Kontextes in Sprachsystemen bereitzustellen. Außerdem werden andere wünschenswerte Merkmale und Charakteristika der vorliegenden Erfindung aus der nachfolgenden detaillierten Beschreibung und den angehängten Ansprüchen offensichtlich, welche in Verbindung mit den beigefügten Zeichnungen und dem vorhergegangenen technischen Bereich und Hintergrund gegeben werden.

[0005] Es werden Verfahren und Systeme für das Steuern eines gesprochenen Dialoges innerhalb eines Sprachsystems bereitgestellt. Das Verfahren beinhaltet das Erstellen einer gesprochenen Dialogsituation, welche einen ersten Dialog-Kontext besitzt, und das Empfangen eines Kontext-Triggers, welcher zu einer Aktion gehört, welche durch den Benutzer durchgeführt wird. In Antwort auf den Kontext-Trigger wechselt das System zu einem zweiten Dialog-Kontext. Nachfolgend kehrt das System in Antwort auf eine Kontext-Vollendungsbedingung dann zu dem ersten Dialog-Kontext zurück.

BESCHREIBUNG DER ZEICHNUNGEN

[0006] Die beispielhaften Ausführungsformen werden hier nachfolgend in Verbindung mit den folgenden gezeichneten Figuren beschrieben, wobei gleiche Ziffern gleiche Elemente bezeichnen und in welchen:

[0007] Fig. 1 ein Funktionsblockdiagramm eines Fahrzeugs ist, welches ein Sprachsystem entsprechend zu verschiedenen beispielhaften Ausführungsformen beinhaltet;

[0008] Fig. 2 ein konzeptionelles Blockdiagramm ist, welches Teilbereiche eines Sprachsystems entsprechend zu verschiedenen beispielhaften Ausführungsformen darstellt;

[0009] Fig. 3 ein Dialog-Kontext-Zustand-Diagramm entsprechend zu verschiedenen beispielhaften Ausführungsformen darstellt; und

[0010] Fig. 4 ein Dialog-Kontext-Verfahren entsprechend zu verschiedenen beispielhaften Ausführungsformen darstellt.

DETAILLIERTE BESCHREIBUNG

[0011] Die folgende detaillierte Beschreibung ist in ihrer Art nur beispielhaft, und es ist nicht beabsichtigt, dass sie die Anmeldung und das Gebrauchsgrenzfeld. Außerdem gibt es keine Absicht, an irgendwelche ausgedrückte oder beinhaltete Theorie, welche in dem vorhergegangenen technischen Bereich, dem Hintergrund, der kurze Zusammenfassung oder der folgenden detaillierten Beschreibung präsentiert wird, gebunden zu sein. Wie es hier gebraucht wird, bezieht sich der Term Modul auf eine anwendungsspezifische integrierte Schaltung (ASIC), eine elektronische Schaltung, einen Prozessor (gemeinsam benutzt, dediziert oder in der Gruppe), einen Speicher, welcher eines oder mehrere Software- oder Firmware-Programme ausführt, eine kombinationslogische Schaltung und/andere geeignete Komponenten.

ten, welche die beschriebene Funktionalität bereitstellen.

[0012] Mit Bezug nun auf **Fig. 1**, entsprechend zu beispielhaften Ausführungsformen des hier beschriebenen Themas, wird ein gesprochenes Dialogsystem (oder einfach ein "Sprachsystem") **10** innerhalb eines Fahrzeugs **12** bereitgestellt. Im Allgemeinen stellt das Sprachsystem **10** Spracherkennung, Dialogsteuerung und Spracherzeugung für ein oder mehrere Fahrzeugsysteme über ein Mensch-Maschine-Schnittstellenmodul-(HMI-)Modul **14** bereit, welches konfiguriert ist, um durch (oder anderenfalls mit einer Schnittstelle) einen oder mehrere Benutzer **40** (z. B. einen Fahrer, einen Passagier, etc.) betrieben zu werden. Derartige Fahrzeugsysteme können zum Beispiel ein Telefonsystem **16**, ein Navigationssystem **18**, ein Mediensystem **20**, ein Telematiksystem **22**, ein Netzsystem **24** und irgendein anderes Fahrzeugsystem beinhalten, welches eine sprachabhängige Anwendung beinhaltet. In einigen Ausführungsformen sind ein oder mehrere der Fahrzeugsysteme kommunikativ an ein Netz (z. B. ein Eigentumsnetz, ein 4 G-Netz oder Ähnliches) gekoppelt, welches Datenkommunikation mit einem oder mehreren Backend-Servern bzw. Nach-Servern **26** bereitstellt.

[0013] Eines oder mehrere mobile Einrichtungen **50** könnten auch innerhalb des Fahrzeugs **12** vorhanden sein, wobei verschiedene Smartphones, Tablet-Rechner, Merkmal-Telefone, etc. beinhaltet sein können. Die mobile Einrichtung **50** kann auch kommunikativ an das HMI **14** über eine gebräuchliche Funkverbindung (z. B. Bluetooth oder WiFi) gekoppelt sein, so dass eine oder mehrere Anwendungen, welche in der mobilen Einrichtung **50** vorhanden sind, für den Benutzer **40** über das HMI **14** zugreifbar sind. Demnach wird ein Benutzer **40** typischerweise Zugriff auf Applikationen besitzen, welche auf drei unterschiedlichen Plattformen laufen: Anwendungen, welche innerhalb der Fahrzeugsysteme selbst ausgeführt werden, Anwendungen, welche auf einer mobilen Einrichtung **50** eingesetzt werden, und Anwendungen, welche auf einem Backend-Server **26** angesiedelt sind. Es wird gewürdigt werden, dass das Sprachsystem **10** in Verbindung mit sowohl auf dem Fahrzeug beruhenden als auch nicht auf dem Fahrzeug beruhenden Systemen benutzt werden kann, welche sprachabhängige Applikationen besitzen, und die auf dem Fahrzeug beruhenden Beispiele, welche hier bereitgestellt sind, werden hier ohne den Verlust der Verallgemeinerung dargelegt.

[0014] Das Sprachsystem **10** kommuniziert mit den Fahrzeugsystemen **14**, **16**, **18**, **20**, **22**, **24** und **26** über einen Kommunikationsbus und/oder ein anderes Datenkommunikationsnetz **29** (z. B. verdrahtet, über Kurzreichweitfunk oder über Langreichweitfunk). Der Kommunikationsbus kann zum Beispiel ein Steuerglied-Bereichsnetz-(CAN-)Bus, ein loka-

ler Zwischenverbindungs-(LIN-)Bus oder Ähnliches sein.

[0015] Wie dargestellt, beinhaltet das Sprachsystem **10** ein Sprachverständnis- bzw. Sprachintelligenzmodul **32**, ein Dialog-Steuermodul **34** und ein Spracherzeugungsmodul **35**. Diese funktionellen Module können als getrennte Systeme oder als ein kombiniertes, integriertes System implementiert sein. Im Allgemeinen empfängt das HMI-Modul **14** ein akustisches Signal (oder eine "Sprachäußerung") **41** von dem Benutzer **40**, welche für das Sprachverständnis-Modul **32** bereitgestellt wird.

[0016] Das Sprachverständnis-Modul **32** beinhaltet jegliche Kombination von Hardware und/oder Software, welche konfiguriert ist, die Sprachäußerung von dem HMI-Modul **14** zu bearbeiten (welche über eine oder mehrere Mikrofone **52** empfangen wird), wobei geeignete Spracherkennungstechniken benutzt werden, welche zum Beispiel automatische Spracherkennung und semantisches Decodieren (oder gesprochenes Sprach-Verständnis (SLU)) beinhalten. Beim Benutzen derartiger Techniken erzeugt das Sprachverständnismodul **32** eine Ergebnisliste (oder einfach eine "Liste") **33** von möglichen Ergebnissen aus der Sprachäußerung. In einer Ausführungsform weist die Liste **33** eine oder mehrere Satzhypothese(n) auf, welche eine Wahrscheinlichkeitsverteilung über den Satz von Äußerungen repräsentieren, welche von dem Benutzer **40** gesprochen sein könnten (d. h. die Äußerung **41**). Die Liste **33** könnte zum Beispiel die Form einer N-Bestenliste haben. In verschiedenen Ausführungsformen erzeugt das Sprachverständnismodul **32** eine Liste **33**, wobei vordefinierte Möglichkeiten benutzt werden, welche in einem Datenspeicher gespeichert sind. Zum Beispiel könnten die vordefinierten Möglichkeiten Namen oder Zahlen sein, welche in einem Telefonbuch gespeichert sind, Namen oder Adressen sein, welche in einem Adressbuch gespeichert sind, Liedernamen, Alben oder Künstler, welche in einem Musikverzeichnis gespeichert sind, etc.. In einer Ausführungsform wendet das Sprachverständnismodul **32** eine Frontend-Merkmalextrahierung aus, welche von einem Hidden-Markov-Modell (HMM) und einem Punktbewertungsmechanismus gefolgt werden.

[0017] Das Dialog-Steuermodul **34** beinhaltet jegliche Kombination von Hardware und/oder Software, welche konfiguriert ist, eine Interaktions- bzw. Wechselwirkungssequenz und eine Auswahl von Sprachkommandos bzw. Sprachaufforderungen **42** zu steuern, welche von dem Benutzer basierend auf der Liste **33** zu sprechen sind. Wenn eine Liste mehr als ein mögliches Ergebnis oder ein Ergebnis mit niedriger Vertrauenswürdigkeit enthält, benutzt das Dialog-Steuermodul **34** Mehrdeutigkeitsstrategien, um eine Interaktion mit dem Benutzer zu steuern, so dass ein erkanntes Ergebnis bestimmt werden kann. Entspre-

chend zu beispielhaften Ausführungsformen ist das Dialog-Steuermodul **34** in der Lage, Dialog-Kontexte zu steuern, wie dies in weiterem Detail nachfolgend beschrieben wird.

[0018] Das Spracherzeugungsmodul **35** beinhaltet jegliche Kombination von Hardware und/oder Software, welche konfiguriert ist, gesprochene Kommandos **42** für einen Benutzer **40** zu erzeugen, basierend auf dem Dialogakt, welcher durch das Dialog-Steuerglied **34** bestimmt ist. Diesbezüglich wird das Spracherzeugungsmodul **35** im Allgemeinen eine natürliche Spracherzeugung (NLG) und eine Sprachsynthese, oder einen Text-in-Sprache (TTS) bereitstellen.

[0019] Die Liste **33** beinhaltet eines oder mehrere Elemente, welche ein mögliches Ergebnis repräsentieren. In verschiedenen Ausführungsformen beinhaltet jedes Element der Liste einen oder mehrere "Slots" bzw. "Positionen", welche jeweils mit einem Slot-Typ bzw. Positionstyp abhängig von der Anwendung verbunden sind. Wenn die Anwendung zum Beispiel das Durchführen von Telefonaten zu Telefonbuchkontakten unterstützt (z. B. "Rufe John Doe an"), dann kann jedes Element Slots mit Slot-Typen eines Vornamens, eines zweiten Vornamens und/oder eines Hauptnamens beinhalten. In einem anderen Beispiel, wenn die Anwendung das Navigieren unterstützt (z. B. "Gehe zu 1111 Sunshine Boulevard"), dann kann jedes Element Slots mit Slot-Typen einer Hausnummer und eines Straßennamens etc. beinhalten. In verschiedenen Ausführungsformen können die Slots und die Slot-Typen in einem Datenspeicher gespeichert werden, und es kann über eines der dargestellten Systeme auf sie zugegriffen werden. Jedes Element oder jeder Slot der Liste **33** ist mit einem Vertrauenspunktwert verbunden.

[0020] Zusätzlich zu dem gesprochenen Dialog könnten Benutzer **40** auch mit dem HMI **14** über verschiedene Tasten, Schalter, Berührungsbildschirm-Benutzerschnittstellenelemente, Gesten (z. B. Handgesten, welche durch eine oder mehrere Kameras erkannt werden, welche innerhalb des Fahrzeugs **12** bereitgestellt sind) und Ähnliches interagieren bzw. Wechselwirken. In einer Ausführungsform wird eine Taste **54** (z. B. eine "Sprechdruck"-Taste oder einfach "Sprechtaste") innerhalb einer leichten Reichweite einem oder mehreren Benutzern **40** bereitgestellt. Zum Beispiel kann die Taste **54** innerhalb eines Lenkrades **56** eingebettet sein.

[0021] Mit Bezug nun auf **Fig. 2**, entsprechend zu verschiedenen beispielhaften Ausführungsformen, beinhaltet das Steuermodul **34** ein Kontext-Handhabungsmodul **202**. Im Allgemeinen beinhaltet das Kontext-Handhabungsmodul **202** jegliche Kombination von Hardware und/oder Software, welche konfiguriert ist, um zu steuern und um zu verstehen, wie Benutzer **40** zwischen den unterschiedlichen Dialog-Kontexten

während einer gesprochenen Dialogsitzung schalten. In einer Ausführungsform beinhaltet zum Beispiel das Kontext-Handhabungsmodul **202** einen Kontext-Stapel **204**, welcher konfiguriert ist, Information zu speichern (z. B. Slot-Information), welche zu einem oder mehreren Dialog-Kontexten gehört, wie dies weiter unten im Detail beschrieben wird.

[0022] Wie es hier benutzt wird, bezieht sich der Term "Dialog-Kontext" im Allgemeinen auf eine spezielle Aufgabe, die ein Benutzer **40** über einen gesprochenen Dialog zu bewerkstelligen versucht, welcher mit einem speziellen Fahrzeugsystem verbunden sein kann oder nicht (z. B. einem Telefonsystem **16** oder einem Navigationssystem **18** in **Fig. 1**). Diesbezüglich können die Dialog-Kontexte visualisiert werden, wie wenn sie eine Baum- oder Hierarchiestruktur besitzen, wobei der Top- bzw. Spitzenknoten der gesamten gesprochenen Dialogsitzung selbst entspricht und die Knoten direkt darunter die allgemeinen Kategorien der Aufgaben aufweisen, welche durch das System bereitgestellt werden, z. B. "Telefon", "Navigation", "Medien", "Klimasteuerung", "Wetter" und Ähnliches. Unter jeden dieser Knoten fallen, spezieller ausgedrückt, Aufgaben, welche zu diesem System gehören. Zum Beispiel kann man unter dem "Navigation"-Knoten neben anderen einen "Wechseln der Navigationseinstellung"-Knoten, einen "Betrachte die Karte"-Knoten und einen "Ziel"-Knoten finden. Unter dem "Ziel"-Knoten kann der Kontextbaum einen "Sehenswürdigkeit"-Knoten, einen "Gebe Adressenknoten ein" und so weiter beinhalten. Die Tiefe und Abmessung eines derartigen Kontextbaumes wird abhängig von der speziellen Anwendung variieren, sie wird im Allgemeinen Knoten am unteren Ende des Baumes beinhalten, welche als "Blatt"-Knoten bezeichnet werden (z. B. Knoten mit keinem weiteren Knoten unterhalb diesem). Zum Beispiel kann das manuelle Eingeben einer speziellen Adresse in das Navigationssystem (und die Zuordnung der dazugehörigen Informations-Slots) als ein Blattknoten betrachtet werden. Im Allgemeinen würden dann die verschiedenen Ausführungsformen, welche hier beschrieben sind, einen Weg für den Benutzer bereitstellen, um sich innerhalb des Kontext-Baumes, welcher durch das Sprachsystem bereitgestellt ist, zu bewegen, und im Speziellen würden sie dem Benutzer gestatten, sich leicht zwischen den Dialog-Kontexten, welche zu den Blattknoten selbst gehören, zu bewegen.

[0023] Mit Bezug nun auf **Fig. 3** (in Verbindung mit beiden **Fig. 1** und **Fig. 2**) kann ein Zustandsdiagramm **300** angewendet werden, um die Art darzustellen, in welcher die Dialog-Kontexte durch das Kontext-Handhabungsmodul **202** gesteuert werden, basierend auf einer Benutzerinteraktion. Im Speziellen repräsentiert der Zustand **302** einen ersten Dialog-Zustand, und der Zustand **304** repräsentiert einen zweiten Dialog-Kontext. Der Übergang **303** vom

Zustand **302** zum Zustand **304** findet in Antwort auf einen "Kontext-Trigger" statt, und der Übergang **305** vom Zustand **304** zum Zustand **302** findet in Antwort auf einen "Kontext-Vollendungszustand" statt. Während **Fig. 3** zwei Dialog-Kontexte darstellt, wird gewürdigt werden, dass einer oder mehrere zusätzliche oder "verschachtelte" Dialog-Kontextzustände während einer speziellen gesprochenen Dialog-Sitzung durchlaufen werden können. Man beachte, dass die Übergänge, welche in dieser Figur dargestellt sind, innerhalb einer einzelnen gesprochenen Dialog-Sitzung stattfinden, anstatt in einer Folge von vielen gesprochenen Dialog-Sitzungen (wenn ein Benutzer eine Sitzung aufgibt und dann in eine andere Sitzung eintritt, um eine nicht bekannte Information zu bestimmen, welche dann in einer nachfolgenden Sitzung benutzt wird).

[0024] Eine große Vielfalt von Kontext-Trigger kann in Verbindung mit dem Übergang **303** benutzt werden. In einem Beispiel wird der Kontext-Trigger gestaltet, um dem Benutzer zu gestatten, leicht und intuitiv zwischen den Dialog-Kontexten zu schalten, ohne einer signifikanten Ablenkung ausgesetzt zu sein. In einer beispielhaften Ausführungsform wird das Aktivieren einer Taste (z. B. einer "Sprechtaste" **54** der **Fig. 1**) als der Kontext-Trigger benutzt. Das heißt, wenn der Benutzer wünscht, die Kontexte zu ändern, drückt der Benutzer einfach die "Sprech"-Taste und fährt mit dem Sprachdialog fort, nun innerhalb eines zweiten Dialog-Kontextes. In einigen Varianten ist die Taste eine virtuelle Taste, z. B. eine Benutzer-Schnittstellenkomponente, welche auf einer zentralen Berührungsschirm-Anzeige bereitgestellt wird.

[0025] In einer anderen Ausführungsform ist der Kontext-Trigger ein vorher ausgewähltes Wort oder eine Phrase, welche durch den Benutzer gesprochen wird – z. B. die Phrase "schalte bzw. wechsele den Kontext". Die vorher ausgewählte Phrase kann vom Benutzer konfigurierbar sein oder kann durch das Kontext-Handhabungsmodul vorher eingestellt sein. Als eine Variation kann ein spezieller Klang (z. B. ein Klick-Geräusch oder ein Pfeifgeräusch, welches durch den Benutzer ausgeführt wird) als der Kontext-Trigger benutzt werden.

[0026] Entsprechend zu einer Ausführungsform wird der Kontext-Trigger in Antwort auf eine natürliche Sprache-Interpretation der Sprache des Benutzers hergestellt, welche vorschlägt, dass der Benutzer wünscht, den Kontext zu ändern. Zum Beispiel kann der Benutzer während einer Navigationssitzung einfach die Phrase sprechen "ich würde gerne nun Jim anrufen, bitte" oder Ähnliches.

[0027] Entsprechend zu einer anderen Ausführungsform wird der Kontext-Trigger in Antwort auf eine Geste erzeugt, welche von einem Benutzer innerhalb des Fahrzeugs durchgeführt wird. Zum Beispiel

sind eine oder mehrere Kameras, welche kommunikativ mit einem Computer-Sichtmodul gekoppelt ist bzw. sind (z. B. innerhalb des HMI **14**), in der Lage, eine Handwelle bzw. ein Handwischen, eine Fingerbewegung oder Ähnliches als gültigen Kontext-Trigger zu erkennen.

[0028] Entsprechend zu einer Ausführungsformen entspricht der Kontext-Trigger dem Sprachsystem **10**, welches erkennt, dass ein unterschiedlicher Benutzer begonnen hat zu sprechen. Das heißt, der Fahrer des Fahrzeugs kann eine gesprochene Dialog-Sitzung initiieren, welche innerhalb eines ersten Dialog-Kontextes stattfindet (z. B. wobei der Fahrer eine Satelliten-Funkstation ändert). Nachfolgend wird, wenn ein Passagier in dem Fahrzeug unterbricht und eine Aufforderung spricht, eine Navigationsaufgabe durchzuführen, der zweite Dialog-Kontext (Navigation zu einer Adresse) eingegeben. Das Sprachsystem **10** kann konfiguriert sein, individuelle Benutzer zu erkennen, wobei eine Vielfalt von Techniken benutzt wird, wobei Sprachanalyse, Richtungsanalyse (z. B. Ort der gesprochenen Stimme) oder ein anderes geeignetes Verfahren benutzt werden.

[0029] Entsprechend einer anderen Ausführungsform entspricht der Kontext-Trigger dem Sprachsystem **10**, wobei bestimmt wird, dass der Benutzer begonnen hat, in einer unterschiedlichen Richtung (z. B. in Richtung eines unterschiedlichen Mikrofons **52**) zu sprechen. Das heißt, der Benutzer kann zum Beispiel einen ersten Dialog-Kontext eingeben, indem er an einem Mikrofon in dem Rückspiegel spricht, und dann den Dialog-Kontext ändern, indem er in ein Mikrofon spricht, welches in der Zentralkonsole eingebettet ist.

[0030] Der Kontext-Vollendungszustand, welcher für den Übergang **305** benutzt wird (d. h. für das Zurückkehren zu dem Originalzustand **302**), kann auch eine Vielfalt von Aktionen umfassen. In einer Ausführungsform zum Beispiel entspricht der Kontext-Vollendungszustand der speziellen Unteraufgabe, welche vollständig ist (z. B. Vollendung eines Telefongesprächs). In einer anderen Ausführungsform kann der Vorgang des erfolgreichen Einfüllens in die erforderlichen "Slots" bzw. „Positionen" der Information selbst, den Kontext-Vollendungszustand umfassen. In einer anderen Weise ausgedrückt, da der Benutzer häufig Dialog-Kontexte zum Zweck des Einfügens fehlender Information schalten bzw. wechseln wird, welche nicht in dem ersten Kontext erlangt ist, kann das System automatisch zu dem ersten Kontext zurückschalten, sobald die erforderliche Information empfangen wird. In anderen Ausführungsformen kann der Benutzer explizit den Wunsch anzeigen, um zu dem ersten Kontext zurückzukehren, wobei zum Beispiel jedes der Verfahren benutzt wird, welches oben in Verbindung mit dem Übergang **303** beschrieben ist.

[0031] Im Folgenden wird ein Beispiel geboten, in welchem ein Benutzer den Kontext ändert, um fehlende Information zu bestimmen, welche der Benutzer dann benutzt, um die Aufgabe zu vollenden:

1. <Benutzer> "Sende Nachricht an John."
2. <System> "Ok. Diktieren Sie eine Nachricht für John."
3. <Benutzer> "Hi, John. Ich bin unterwegs, und ich werde... da sein"
4. <Benutzer> [aktiviert Kontext-Trigger]
5. <Benutzer> "Wann ist meine Ankunftszeit?"
6. <System> "Ihre geschätzte Ankunftszeit ist vier Uhr nachmittags."
7. <Benutzer> ungefähr vier Uhr nachmittags"

[0032] Wie aus diesem Beispiel ersehen werden kann, wird der erste Dialog-Kontext (welcher eine Sprachnachricht verfasst) durch den Benutzer im Schritt 4 unterbrochen, um die geschätzte Zeit während eines zweiten Dialog-Kontextes zu bestimmen (eine Navigations-Vollendungsabschätzung). Nachdem das System die geschätzte Ankunftszeit bereitstellt, kehrt das System automatisch zu dem ersten Dialog-Kontext zurück. Das vorherige Diktieren wurde ungeachtet des Dialog-Kontextschalters aufrechterhalten, und demnach kann der Benutzer einfach mit der diktierten Nachricht fortfahren, welche dort startet, wo er sie verlassen hat.

[0033] Im Folgenden wird ein weiteres Beispiel präsentiert, in welchem die Information, welche der Benutzer für einen inkorrekten Dialogpfad, welcher von dem System hergenommen wird, korrigiert.

1. <Benutzer> "Spiele John Lennon."
2. <System> "Ok. Stelle Ziel auf John Lennon-Avenue bzw. -Straße ein. Bitte gebe Nummer ein."
3. <Benutzer> "Warte. Ich möchte Musik hören."
4. <System> "Ok. Welches Album oder welchen Titel?"

[0034] In dem obigen Beispiel hat das System bei Schritt 2 die Sprache des Benutzers missverstanden und hat einen Navigations-Dialog-Kontext eingegeben. Der Benutzer benutzt dann eine vorher festgelegte Phrase "Warte" als einen Kontext-Schalter, was das System veranlasst, einen Medien-Dialog-Kontext einzugeben. Alternativ kann das System "Warte. Ich wünsche Musik zu hören" mit Hilfe der natürlichen Sprachanalyse interpretiert haben, um auf die Absicht des Benutzers rückzuschließen.

[0035] Das folgende Beispiel ist auch erläuternd für einen Fall, in welchem der Benutzer von einem Navigations-Dialog-Kontext zu einem Telefonanruf-Kontext wechselt, um die fehlende Information zu bestimmen.

1. <Benutzer> "Finde für mich ein Restaurant, welches Meeresfrüchte anbietet."
2. <System> "Bill's Krabbenbude ist eine halbe Meile weg und bietet Meeresfrüchte an."

3. <Benutzer> "Welches ist deren Preisbereich?"
4. <System> "Tut mir Leid. Keine Preisbereichsinformation verfügbar."
5. <Benutzer> [aktiviert Kontext-Trigger]
6. <Benutzer> "Rufe Bob an."
7. <System> "Rufe Bob an."
8. <Bob> "Hallo?"
9. <Benutzer> "Hallo, Bob. Ist Bill's Krabbenbude teuer?"
10. <Bob> "Hm, nein. Es ist eine 'Krabbenbude'."
11. <Benutzer> "Danke. Tschüss." [hängt auf]
12. <Benutzer> "Ok. Bitte führe mich dorthin."
13. <System> "Bin dabei, Ziel zu laden..."

[0036] In anderen Ausführungsformen wird die fehlende Information von dem zweiten Dialog-Kontext zurück zu dem ersten Dialog-Kontext automatisch beim Zurückkehren übertragen.

[0037] Mit Bezug nun auf das Ablaufdiagramm, welches in Fig. 4 in Verbindung mit Fig. 1–Fig. 3 dargestellt wird, wird nun ein beispielhaftes Kontext-Schaltverfahren 400 beschrieben. Es sollte beachtet werden, dass das dargestellte Verfahren nicht auf die Folge begrenzt ist, welche in Fig. 4 gezeigt wird, sondern in einer oder mehreren variierenden Reihenfolgen, soweit anwendbar, durchgeführt werden kann. Außerdem können ein oder mehrere Schritte des dargestellten Verfahrens in verschiedenen Ausführungsformen hinzugefügt oder weggelassen werden.

[0038] Anfangs wird angenommen, dass eine gesprochene Dialog-Sitzung erstellt worden ist und dass sie entsprechend einem ersten Dialog-Kontext fortfährt. Während dieser Sitzung aktiviert der Benutzer den richtigen Kontext-Trigger (402), wie zum Beispiel einen der Kontext-Trigger, welche oben beschrieben sind. In Antwort darauf drückt das Kontext-Steuermodul 202 auf den Kontext-Stapel 204, den aktuellen Kontext (404) und die Rückkehradresse (406). Das heißt, der Kontext-Stapel 204 weist einen Zuerst-ein-, zuletzt-aus-(FILO-)Stapel auf, welcher Information bezüglich eines oder mehrerer Dialog-Kontexte speichert. Ein "Drücken" platziert einen Begriff auf dem Stapel, und ein "Pop" entfernt einen Begriff von dem Stapel. Die gedrückte Information wird typischerweise Daten beinhalten (z. B. "Slot-Information"), welche zu der Aufgabe gehören, welche in diesem speziellen Kontext durchgeführt werden. Fachleute werden erkennen, dass der Kontext-Stapel 204 in vielfältiger Weise implementiert werden kann. In einer Ausführungsform wird zum Beispiel jeder Dialog-Status als eine Klasse implementiert und ist ein Knoten in einem Dialog-Baum, wie oben beschrieben. Die Phrasen "Klasse" und "Gegenstand" werden hier übereinstimmend mit ihrem Gebrauch in Verbindung mit gewöhnlichen objektorientierten Programmiersprachen gebraucht, wie zum Beispiel Java oder C++. Die Rückadresse entspricht dann einem Zeiger für die Kontext-Instantiierung bzw. -Umschrei-

bung. Die vorliegende Offenbarung ist jedoch nicht so begrenzt und kann implementiert werden, indem eine Vielzahl von Programmiersprachen benutzt wird.

[0039] Als Nächstes schaltet in **408** das Kontext-Handhabungsmodul **202** zu der Adresse, welche dem zweiten Kontext entspricht. Beim Eintreten des zweiten Kontextes wird eine Bestimmung durchgeführt, ob das System in diesem Kontext als Teil eines "Schaltens" von einem anderen Kontext eingetreten ist (**410**). Wenn dem so ist, fährt der gesprochene Dialog fort, bis der Kontext-Vollendungszustand aufgetreten ist (**412**), worauf die Ergebnisse des zweiten Kontextes selbst auf den Kontextstapel **204** gedrückt werden (**414**). Als Nächstes stellt das System die (zuvor gedrückte) Rückadresse von dem Kontextstapel **204** wieder her und kehrt zu dem ersten Dialog-Kontext zurück (**416**). Als Nächstes werden innerhalb des ersten Dialog-Kontextes die Ergebnisse (von dem zweiten Dialog-Kontext) von dem Kontextstapel **204** gelesen (**418**). Der Original-Dialog-Kontext, welcher auf den Kontextstapel **204** während **404** gedrückt wurde, wird dann wiederhergestellt und in den ersten Dialog-Kontext eingearbeitet (**420**). Auf diese Weise können die Dialog-Kontexte mitten in der Sitzung geschaltet bzw. gewechselt werden, ohne von dem Benutzer zu erfordern, eine erste Sitzung zu beenden, eine neue Sitzung zu starten, um die fehlende Information (oder Ähnliches) zu bestimmen und dann noch eine weitere Sitzung zu beginnen, um die Aufgabe zu beenden, welche ursprünglich für die erste Sitzung beabsichtigt war. In anderer Weise ausgedrückt, ein Satz von Daten, welcher während des zweiten Dialog-Kontextes bestimmt ist, wird optional in einen anderen Satz von Daten eingearbeitet, welche während des ersten Dialog-Kontextes bestimmt sind, um eine Sitzungsaufgabe zu erfüllen.

[0040] Während wenigstens eine beispielhafte Ausführungsform in der vorhergegangenen detaillierten Beschreibung präsentiert wurde, sollte gewürdigt werden, dass eine große Anzahl von Variationen existiert. Es sollte auch gewürdigt werden, dass die beispielhafte Ausführungsform oder die beispielhaften Ausführungsformen nur Beispiele sind und es nicht beabsichtigt ist, dass diese den Umfang, die Anwendbarkeit oder die Konfiguration der Offenbarung in irgendeiner Weise begrenzen. Vielmehr wird die vorhergegangene detaillierte Beschreibung Fachleuten eine bequeme Anleitung für das Implementieren der beispielhaften Ausführungsform oder der beispielhaften Ausführungsformen bereitstellen. Es sollte davon ausgegangen werden, dass verschiedene Änderungen in der Funktion und in der Anordnung der Elemente durchgeführt werden können, ohne vom Umfang der Offenbarung abzuweichen, wie er in den angehängten Ansprüchen und den rechtlichen Äquivalenten davon dargelegt ist.

WEITERE AUSFÜHRUNGSFORMEN

1. Verfahren für das Steuern eines gesprochenen Dialoges innerhalb eines Sprachsystems, wobei das Verfahren aufweist:
 - Erstellen einer gesprochenen Dialog-Sitzung, welche einen ersten Dialog-Kontext besitzt;
 - Empfangen eines Kontext-Triggers, welcher zu einer Aktion gehört, welche von einem Benutzer durchgeführt wird;
 - in Antwort auf den Kontext-Trigger, Wechseln zu einem zweiten Dialog-Kontext; und
 - in Antwort auf einen Kontext-Vollendungszustand, Zurückkehren zu dem ersten Dialog-Kontext.
2. Verfahren nach Ausführungsform 1, wobei die Aktion, welche von dem Benutzer durchgeführt wird, einem Drücken einer Taste entspricht.
3. Verfahren nach Ausführungsform 2, wobei das Drücken der Taste dem Drücken einer Taste entspricht, welche in einem Lenkrad eines Automobils eingebaut ist.
4. Verfahren nach Ausführungsform 1, wobei die Aktion, welche von dem Benutzer ausgeführt wird, wenigstens einem von folgenden entspricht: Sprechen einer vorher ausgewählten Phrase, Durchführen einer Geste und Sprechen in eine vorher festgelegte Richtung.
5. Verfahren nach Ausführungsform 1, wobei die Daten, welche während des zweiten Dialog-Kontextes bestimmt sind, in Daten eingearbeitet werden, welche während des ersten Dialog-Kontextes bestimmt sind, um eine Sitzungsaufgabe durchzuführen.
6. Verfahren nach Ausführungsform 5, welches ferner das Drücken des zweiten Satzes von Daten auf einen Kontextstapel, vor dem Wechseln zu dem zweiten Dialog-Kontext, aufweist.
7. Sprachsystem, welches aufweist:
 - ein Sprache-Verständnis-Modul, welches konfiguriert ist,
 - eine Sprachäußerung von einem Benutzer zu empfangen und eine Ergebnisliste bereitzustellen, welche zu der Sprachäußerung gehört;
 - ein Dialog-Steuermodul, welches kommunikativ an das Sprache-Verständnis-Modul gekoppelt ist, wobei das Dialog-Steuermodul ein Kontext-Handhabungs-Modul beinhaltet, welches konfiguriert ist, um: die Ergebnisliste zu empfangen;
 - mit dem Benutzer eine gesprochene Dialogsitzung zu erstellen, welche einen ersten Dialogkontext besitzt, basierend auf der Ergebnisliste;
 - einen Kontext-Trigger zu empfangen, welcher zu einer Aktion gehört, welche durch einen Benutzer durchgeführt wird;
 - in Antwort auf den Kontext-Trigger, zu einem zweiten Dialog-Kontext zu wechseln; und
 - in Antwort auf einen Kontext-Vollendungszustand, zu dem ersten Dialog-Kontext zurückzukehren.

8. Sprachsystem nach Ausführungsform 7, wobei der Kontext-Trigger das Drücken einer Taste aufweist.

9. Sprachsystem nach Ausführungsform 8, wobei das Drücken der Taste dem Drücken einer Taste entspricht, welche in einem Lenkrad eines Automobils eingebaut ist.

10. Sprachsystem nach Ausführungsform 7, wobei der Kontext-Trigger eine vorher ausgewählte Phrase aufweist, welche von dem Benutzer gesprochen ist.

11. Sprachsystem nach Ausführungsform 7, wobei der Kontext-Trigger eine Geste aufweist, welche von dem Benutzer durchgeführt wird.

12. Sprachsystem nach Ausführungsform 7, wobei der Kontext-Trigger eine Bestimmung aufweist, dass der Benutzer in eine vorher festgelegte Richtung spricht.

13. Sprachsystem nach Ausführungsform 7, wobei der Kontext-Trigger eine Bestimmung aufweist, dass ein zweiter Benutzer begonnen hat, zu sprechen.

14. Sprachsystem nach Ausführungsform 7, wobei die Daten, welche während des zweiten Dialog-Kontextes bestimmt sind, in Daten eingearbeitet sind, welche während des ersten Dialog-Kontextes bestimmt sind, um eine Sitzungsaufgabe durchzuführen.

15. Sprachsystem nach Ausführungsform 14, wobei das Kontext-Handhabungsmodul einen Kontext-Stapel beinhaltet und konfiguriert ist, den zweiten Satz von Daten auf den Kontext-Stapel vor dem Wechseln zu dem zweiten Dialog-Kontext zu drücken.

16. Sprachsystem nach Ausführungsform 7, wobei der Kontext-Vollendungsstatus die Vollendung einer Unteraufgabe aufweist, welche von dem Benutzer durchgeführt wird.

17. Nicht-transitorisches, von einem Computer lesbares Medium, welches Software-Instruktionen trägt, wobei die Software-Instruktionen konfiguriert sind, ein Sprachsystem zu instruieren, um: mit einem Benutzer eine gesprochene Dialog-Sitzung zu erstellen, welche einen ersten Dialog-Kontext besitzt;

einen Kontext-Trigger zu empfangen, welcher zu einer Aktion gehört, welche von einem Benutzer durchgeführt wird;

in Antwort auf den Kontext-Trigger, Wechseln zu einem zweiten Dialog-Kontext; und

in Antwort auf einen Kontext-Vollendungsstatus, Zurückkehren zu dem ersten Dialog-Kontext.

18. Nicht-transitorisches, von einem Computer lesbares Medium nach Ausführungsform 17, wobei der Kontext-Trigger dem Drücken einer Taste entspricht, welche in einem Lenkrad eines Automobils eingebaut ist.

19. Nicht-transitorisches, von einem Computer lesbares Medium nach Ausführungsform 17, wobei die Daten, welche während des zweiten Dia-

log-Kontextes bestimmt sind, in Daten eingearbeitet werden, welche während des ersten Dialog-Kontextes bestimmt sind, um eine Sitzungsaufgabe durchzuführen.

20. Nicht-transitorisches, von einem Computer lesbares Medium nach Ausführungsform 19, wobei die Software-Instruktionen den Prozessor instruieren, den zweiten Satz von Daten auf einen Kontext-Stapel zu drücken, vor dem Wechseln zu dem zweiten Dialog-Kontext.

Patentansprüche

1. Verfahren für das Steuern eines gesprochenen Dialoges innerhalb eines Sprachsystems, wobei das Verfahren aufweist:

Erstellen einer gesprochenen Dialog-Sitzung, welche einen ersten Dialog-Kontext besitzt;

Empfangen eines Kontext-Triggers, welcher zu einer Aktion gehört, welche von einem Benutzer durchgeführt wird;

in Antwort auf den Kontext-Trigger, Wechseln zu einem zweiten Dialog-Kontext; und

in Antwort auf einen Kontext-Vollendungsstatus, Zurückkehren zu dem ersten Dialog-Kontext.

2. Verfahren nach Anspruch 1, wobei die Aktion, welche von dem Benutzer durchgeführt wird, einem Drücken einer Taste entspricht.

3. Verfahren nach Anspruch 2, wobei das Drücken der Taste dem Drücken einer Taste entspricht, welche in einem Lenkrad eines Automobils eingebaut ist.

4. Verfahren nach einem der vorherigen Ansprüche, wobei die Aktion, welche von dem Benutzer ausgeführt wird, wenigstens einem von folgenden entspricht: Sprechen einer vorher ausgewählten Phrase, Durchführen einer Geste und Sprechen in eine vorher festgelegte Richtung.

5. Verfahren nach einem der vorherigen Ansprüche, wobei die Daten, welche während des zweiten Dialog-Kontextes bestimmt sind, in Daten eingearbeitet werden, welche während des ersten Dialog-Kontextes bestimmt sind, um eine Sitzungsaufgabe durchzuführen.

6. Verfahren nach einem der vorherigen Ansprüche, welches ferner das Drücken des zweiten Satzes von Daten auf einen Kontext-Stapel, vor dem Wechseln zu dem zweiten Dialog-Kontext, aufweist.

7. Sprachsystem, welches aufweist: ein Sprache-Verständnis-Modul, welches konfiguriert ist, eine Sprachäußerung von einem Benutzer zu empfangen und eine Ergebnisliste bereitzustellen, welche zu der Sprachäußerung gehört; ein Dialog-Steuermodul, welches kommunikativ an das Sprache-Verständnis-Modul gekoppelt ist, wobei

das Dialog-Steuermodul ein Kontext-Handhabungs-Modul beinhaltet, welches konfiguriert ist, um: die Ergebnisliste zu empfangen; mit dem Benutzer eine gesprochene Dialogsitzung zu erstellen, welche einen ersten Dialogkontext besitzt, basierend auf der Ergebnisliste; einen Kontext-Trigger zu empfangen, welcher zu einer Aktion gehört, welche durch einen Benutzer durchgeführt wird; in Antwort auf den Kontext-Trigger, zu einem zweiten Dialog-Kontext zu wechseln; und in Antwort auf einen Kontext-Vollendungszustand, zu dem ersten Dialog-Kontext zurückzukehren.

8. Sprachsystem nach Anspruch 7, wobei der Kontext-Trigger das Drücken einer Taste aufweist.

9. Sprachsystem nach Anspruch 8, wobei das Drücken der Taste dem Drücken einer Taste entspricht, welche in einem Lenkrad eines Automobils eingebaut ist.

10. Sprachsystem nach Anspruch 7, wobei der Kontext-Trigger eine vorher ausgewählte Phrase aufweist, welche von dem Benutzer gesprochen ist, eine Geste aufweist, welche von dem Benutzer durchgeführt wird, eine Bestimmung aufweist, dass der Benutzer in eine vorher festgelegte Richtung spricht und/oder eine Bestimmung aufweist, dass ein zweiter Benutzer begonnen hat, zu sprechen.

Es folgen 3 Seiten Zeichnungen

Anhängende Zeichnungen

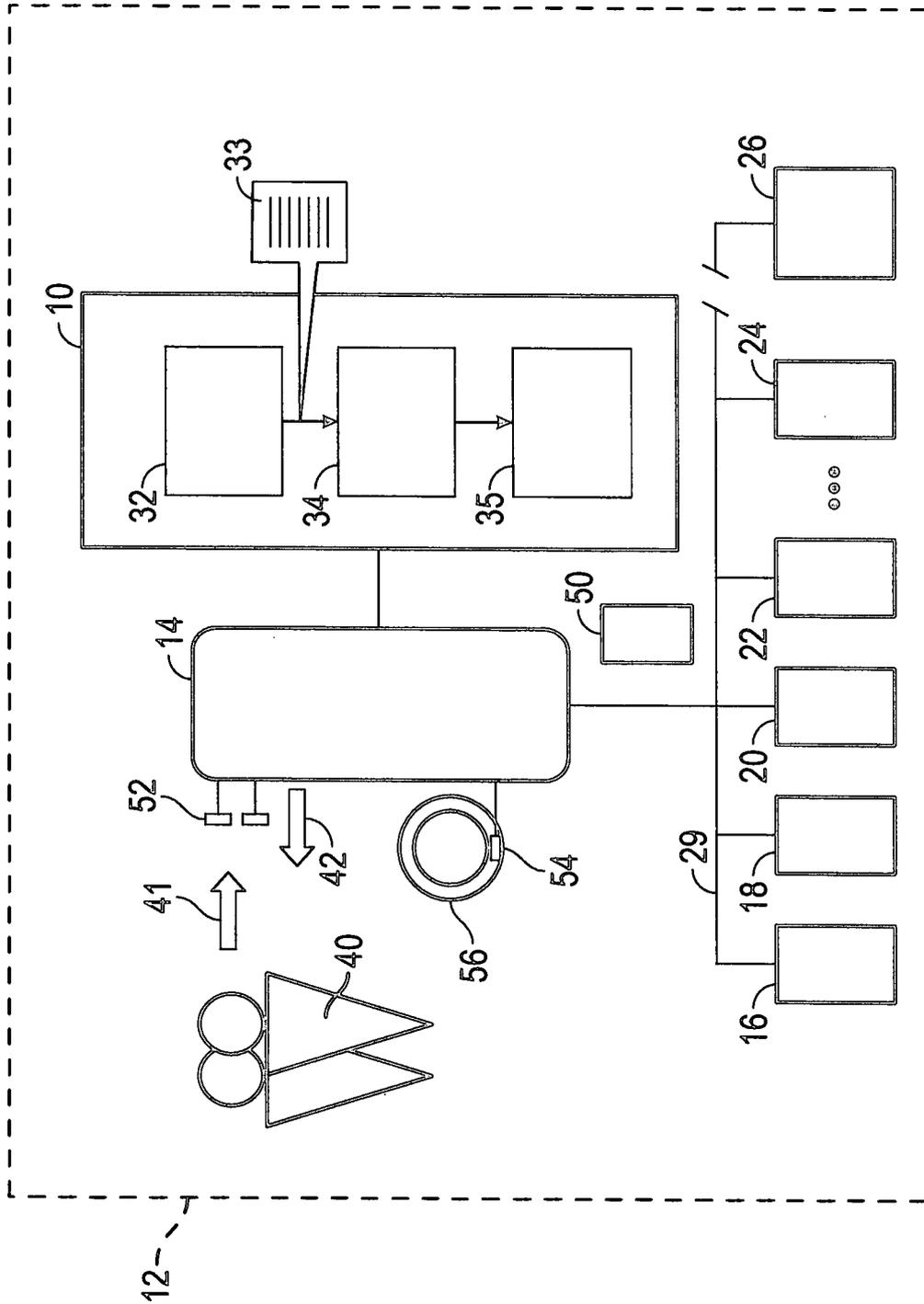


FIG. 1

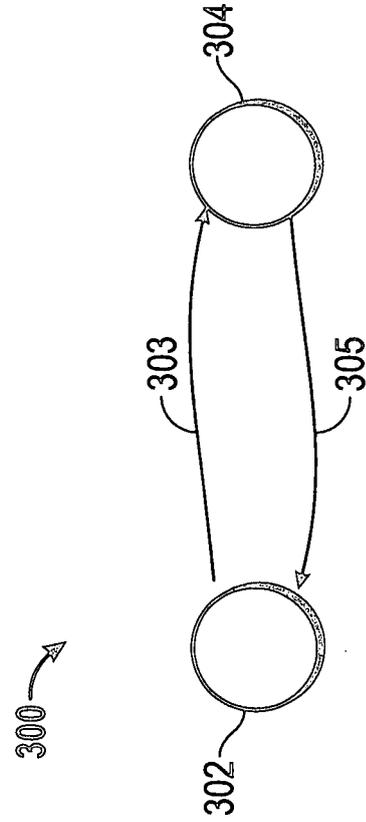


FIG. 3

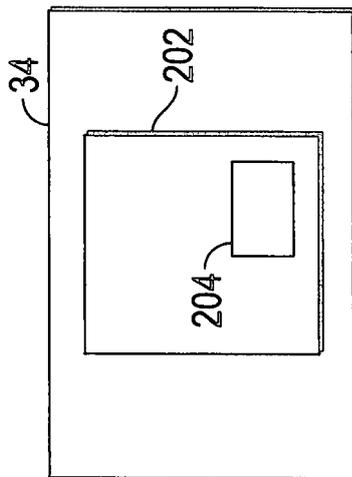


FIG. 2

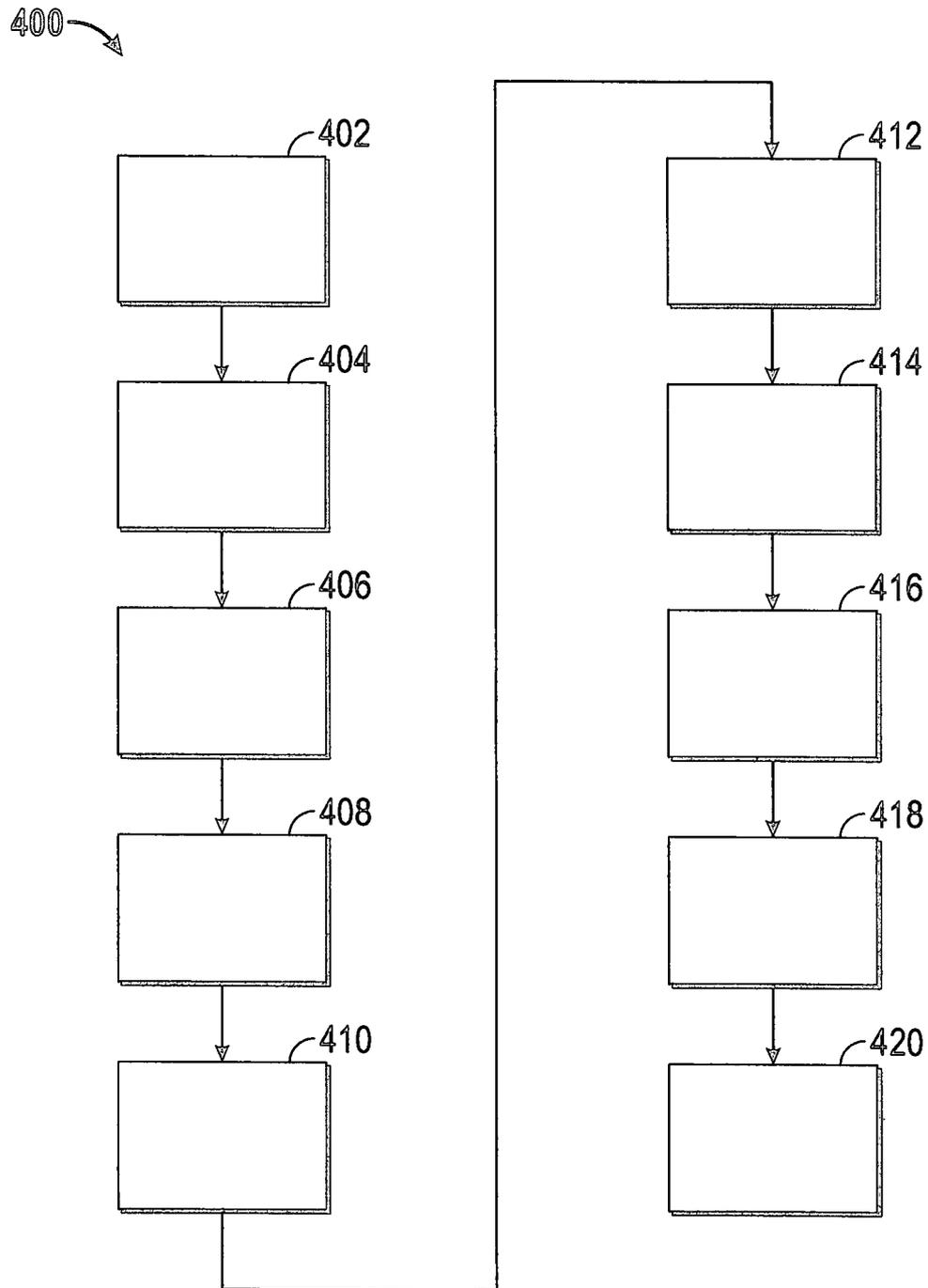


FIG. 4